MASTER'S THESIS

Agile facilitates data-driven solutions

Donker - Rostamy, M. (Maryam)

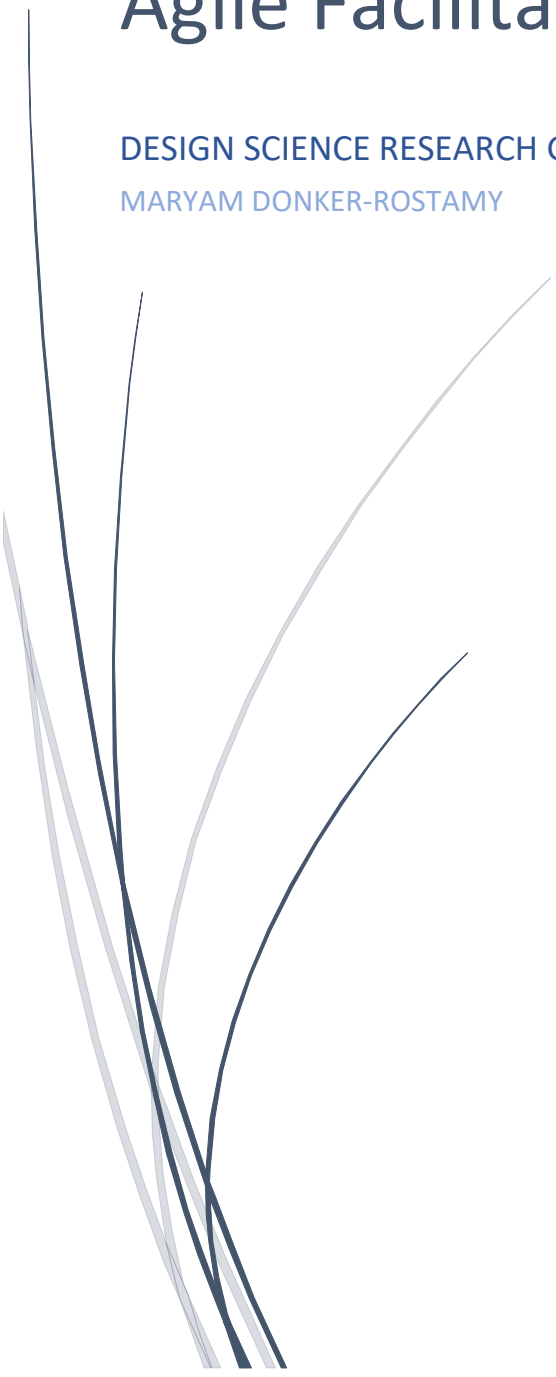
Award date:
2019

2019

Agile Facilitates Data-Driven Solutions

DESIGN SCIENCE RESEARCH GRADUATION PROJECT — THESIS

MARYAM DONKER-ROSTAMY



"Intelligence is the ability to adapt to change."
- Stephen Hawking

“Agile Facilitates Data-Driven Solutions” Framework in Data-Driven Projects

Degree program:	Open University of the Netherlands, Faculty of Management, Science and Technology Business Process Management and IT master's program
Course:	IM0602 BPMIT Graduation Assignment Preparation IM9806 Business Process Management and IT Graduation Assignment
Student:	Maryam Donker-Rostamy
Identification number:	
Date:	Wednesday, July 2019
Thesis supervisor	Prof.dr.ir. Remko Helms
Second reader	Jeroen Baijens Msc. 1.0.0.
Version number:	Final
Status:	

Abstract

This design science research presents an Agile framework that guides organizations in adopting Agile in their data-driven projects. The framework consists of guidelines that support the organization in 1) beginning to implement Agile via minor changes, 2) selecting the applicable Agile principles and methodologies for their projects, and 3) applying these methodologies and principles to improve the development process and performance of data-driven teams. For the construction and rigorous evaluation of the framework, this research utilizes expert interviews, online questionnaire presentations, and a workshop. The research is based on the knowledge gained from prior research, and the requirements of the framework were gathered based on the theoretical framework and validated in a business environment. The results of this research contribute to the design and development of relevant solutions for adopting Agile in the field of data product development.

Key terms

Agile Data-Driven Framework, Agile Data Development Model, Agile Principles, -Practices, and -Methodologies

Summary

The literature review of this paper demonstrates that implementing Agile in data-driven projects enhances the process of data development. This study determines that the four values of the Agile Manifesto and the twelve principles of Agile must be completely understood by project team participants and must be adopted in every organization that attempts to implement Agile. Furthermore, an iterative and incremental development process, together with Agile practices, can support data-driven projects and can improve the performance of teams.

The challenge for data-driven projects, is to adopt Agile practices and methodologies that most effectively facilitate these projects. To address this problem, the design science research (DSR) framework of Hevner et al. (2004) is used as a research methodology for this thesis to design and develop a framework as the solution to the business problem. The DSR of this thesis answers the following main research question:

How can Agile principles, practices, and methodologies be applied in data-driven projects to improve and effectively facilitate these projects?

This design science research, presents "Agile Facilitates Data-Driven Solutions" (AFDS) framework as the designed artifact of the DSR which supports data-driven organizations in applying Agile in their projects. The framework is designed, developed, demonstrated and evaluated within a Dutch data-driven consulting company.

The AFDS framework targets each member of a project organization and helps data-driven teams create small, shippable data products and actionable insights in short iterations. The framework suggests a development cycle that is extracted from the fast analytics/data science lifecycle (Larson & Chang, 2016); the framework also adopts some of the Scrum guidelines to provide an iterative and incremental development method that develops data products in short iterations. Furthermore, the framework proposes a set of Agile practices that can be used by data-driven multidisciplinary teams to improve their performance.

The requirements of the framework gathered from the literature review are validated by an online questionnaire and expert interviews. Once the draft version of the design was ready, the framework was presented to the experts who were interviewed. The feedback of the experts was incorporated into the design to improve the framework. Finally, when the design was complete, due to time limitation, the framework was demonstrated and evaluated in the form of a presentation followed by a workshop and an online survey at the company. The results indicate that the AFDS framework is appropriate for data-driven projects. The framework is easy to use and understand and provides sufficient guidelines. However, the results of the evaluation (i.e., the presentation and workshop) indicate that a lack of knowledge about data development cycles and a lack of experience in using Agile can cause issues that could interfere with the implementation of the framework in projects. These issues could create difficulties for the company when applying the framework in its projects. To address these issues, a qualitative study should be performed to test the framework in real-world environments and to prove the applicability and completeness of the framework.

It is also recommended that the company provide training, workshops, and meet-ups for its teams to improve their knowledge and experiences about data development approaches and lead them to adopt Agile in their daily work. These actions would allow the project and its teams to learn and develop a better understanding of the effects of applying Agile and using the Agile framework in their projects. If team members are more aware of the content of the framework and know how to apply it in their daily work, they could share their experiences more convincingly with others and suggest that they apply it too.

Table of Contents

Table of Figures

Table of tables

1. Introduction

1.1. Background

The 21st century has introduced the age of data, data analytics, data science, and big data (Cao, 2017). The global competition in knowledge-based economies is based on transforming data into knowledge and information (Nascimento, Santana, & de Oliveira, 2012). Anderson (2015) defines data-drivenness as "*building tools, abilities, and, most crucially, a culture that acts on data*." Dhar (2012) and Jurney (2017) consider data science as an interdisciplinary field that extracts knowledge and actionable insights from data by using scientific algorithms and methods. According to Saltz et al. (2018), data science emerged from statistics, computer science, and information management as a distinct discipline that includes the collection, preparation, modeling, evaluation, deployment, and dissemination of data. By using data, organizations are better able to intelligently sell their products and services and become stronger in the knowledge-based economy. Since 1990, several approaches to data-driven projects and the development of data products have been defined. Figure 1 illustrates the evolution of data process models. Mariscal et al. (2010) divide these process models into the following three general approaches:

1. Approaches related to KDD (Knowledge Discovery in Data)
2. Approaches related to CRISP-DM (Cross Industry Standard Process for Data Mining)
3. Other independent approaches, such as KDD Roadmap

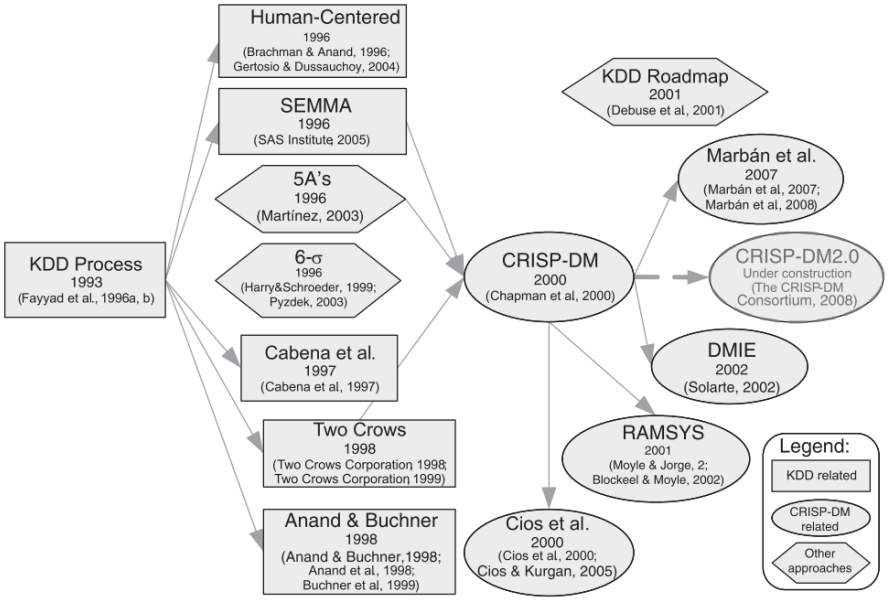


Figure 1: Evolution of data mining process models (Mariscal et al., 2010)

The logical nature of these approaches is based on linear and sequential processes and is established via the Waterfall software development method (Schmidt & Sun, 2018). The Waterfall method is restrictive and inflexible, which indicates the entire solution must be completed before it can be launched and deliver value to a business. Waterfall is plan-driven, which implies that the scope, budget, and deadlines are established in advance, and projects attempt to avoid changes (Larson & Chang, 2016). According to Asay (2017), "85% of the big data projects fail […]." As business requirements become more dynamic and uncertain, the traditional plan-driven and linear approaches are unable to follow the dynamic requirement changes in a rapidly evolving environment (Nascimento et al., 2012).

In the 19th century, most software projects exceeded their budgets and were behind schedule. The lack of productivity and quality were considered the source of this crisis. Therefore, in February 2001, Kent Beck, alongside software gurus and leaders of XP programming, introduced the Agile way of working and published the "Manifesto of Agile Software Development". Agile means viable. An Agile software development team fully understands the manifesto of Agile and applies it in its daily work. The goal of the team is to iteratively and incrementally deliver a minimum viable product (MVP) to its customer.

It is evident from studies that the popular Agile revolution in software development affected the young and rapidly growing field of data product development. Agile seems to be a possible solution to problems caused by rapidly and frequently changing requirements. Furthermore, Agile increase the success rate of data-driven projects. This thesis proposes an Agile framework that provides guidelines that support organizations in applying Agile to improve their development processes and increase the success of their data-driven projects.

1.2. Problem statement

An increasing number of organizations are struggling to manage their data-driven projects. These projects must deliver faster and better data products. Step-by-step linear and sequential approaches such as KDD and CRISP-DM focus on "what to do" when developing a data product and not "how to do it." The descriptions of these approaches provide no guidance on how data-driven projects should work to successfully develop a product in an business environment in which the requirements are rapidly and continuously changing (Saltz, Shamshurin, & Croswston, 2017).

Schmidt and Sun (2018) reveal that Agile data-driven projects can build predictive models, which could be continuously refined and improved; however, it appears that these projects do not yet have enough experience with Agile to understand how to adapt it to their situations. Thus, the problem can be described as follows:

The linear and traditional approaches are the reasons that data-driven projects continue to fail. Unfortunately, a lack of knowledge and experience about Agile in data-driven organizations causes them to encounter daunting challenges in adopting the Agile way of working and applying Agile practices and methodologies to their projects.

1.3. Research objective and questions

Based on its adaptive nature, Agile should be suitable for nearly any project; however, this requires a deeper understanding of its values and principles. If organizations manage the inherent uncertainty of Agile and have the courage to slowly approach an Agile transition, the great challenge remains of adopting the practices and methodologies that most effectively facilitate their data-driven projects. Therefore, the research questions for this thesis are:

Main question:

How can Agile principles, practices, and methodologies be applied in data-driven projects to improve and effectively facilitate these projects?

The main question is divided into the following sub-questions:

Sub question 1:

What is the development cycle of a data-driven project?

This first sub-question is derived from observations of the data development cycle and the fundamental characteristics of data as a product. This sub-question is answered by the literature review.

Sub question 2*:*

Which Agile principles, practices, and methodologies can be applied in data-driven projects?

This sub-question is determined by identifying the Agile principles, practices, and methodologies that can be used in data-driven projects. This sub-question is also answered by the literature review.

Once the Agile principles, practices, and methodologies and their impact on the data-driven organizations are studied, the design science research (DSR) of this thesis begins to answer the main question above.

1.4. Motivation/relevance

It has become apparent from various scientific case studies that Agile is applied in data-driven projects. To date, however, there is a lack of design science theories that can support data-driven organizations in adopting Agile to improve their projects. To address this dilemma, this thesis identifies and structures three contributions to the field of information science.
First, the literature review of this thesis considers the relevant scientific literature. The review analyzes the research questions and proposes a theoretical overview, which is presented in Figure 5. This theoretical overview defines the scope of the DSR, which is based on prior knowledge, and determines the properties, components, and design requirements of the DSR artifact. Second, throughout this DSR, an Agile framework is designed and evaluated in a data-driven organization that desires to transition from traditional methods to a dynamic Agile project approach. The framework addresses unresolved issues related to adopting Agile in data-driven projects and provides answers to the main research question. Finally, the results of this DSR support the organization in adopting Agile to improve its data-driven projects.

1.5. Research approach

DSR is selected as the research approach to discover, identify, and provide the ideal opportunities to answer the research questions. Figure 2 below represents the activities that are depicted in the design science research methodology (DSRM) introduced by Peffers et al. (2007).

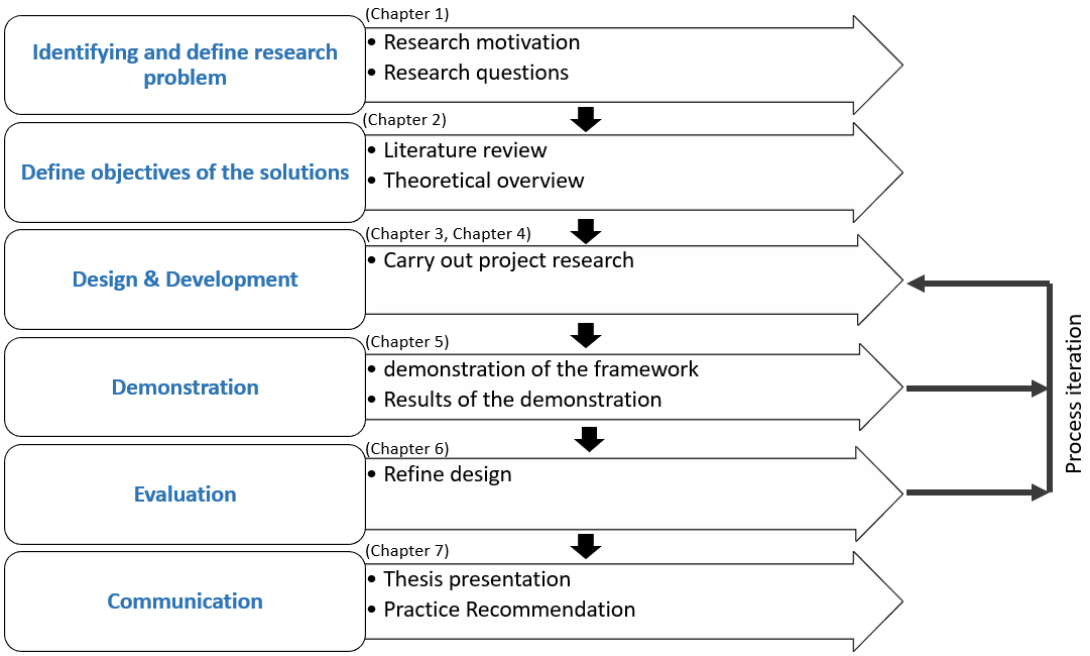


Figure 2: Research approach (Peffers et al., 2007)

The research begins by identifying the problem statement and research questions. Afterward, the literature review is performed to review prior literature. A theoretical framework is represented as a result of the literature review, which presents the previous literature and theories concerning Agile values, principles, practices, and methodologies. The literature review answers the two sub-questions of this research. The "Design and Development" step of the DSRM employs the theoretical overview to create an artifact for answering the main question of the research.

The framework is designed, developed, and evaluated within a data-driven organization in the Netherlands. First, the observation and validation of the requirements is conducted using a questionnaire and five semi-structured expert interviews. The outcome of the questionnaire and expert interviews led to the design of the framework. The design of the framework is formative evaluated using a presentation. During the presentation, the framework is presented to the experts who were interviewed. When new requirements were identified by the evaluation, the process returned to the design and development step to enhance the framework.
The framework is demonstrated in a presentation. The demonstration is limited to a presentation due to time constraints. The demonstration is evaluated through a workshop and an online survey. The demonstration is evaluated through a workshop and an online survey. Finally, the findings and outcomes obtained throughout the research are presented using a thesis presentation and defense (Peffers et al., 2007; Saunders, Lewis, & Thornhill, 2012). Further explanation of DSR and DSRM is included in chapter 3.

2. Literature review

The literature review discusses the existing theory concerning the research problem and determines information established in prior literature. This chapter presents the approach, implementation, and results of the literature review.

2.1. Approach

Several aspects of the guideline "Systematic Literature Review (SLR) of Information Systems Research (ISR)" introduced by Okoli and Schabram (2011) are employed to conduct the literature review. Figure 3 displays the SLR guideline and its steps.

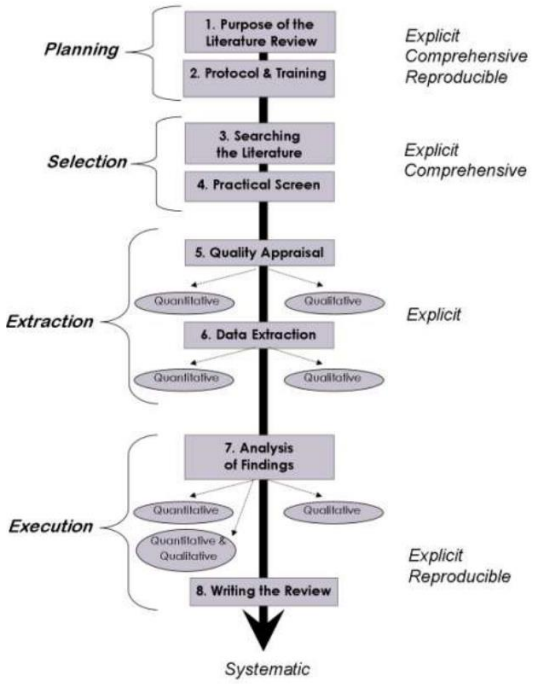


Figure 3: Systematic literature review (Okoli & Schabram, 2011)

This literature review aims to answer the research sub-questions presented in section 1.3. The results of this study are formulated in a theoretical overview that supports the design and development of the research artifact further in the research process.

Once the purpose of the study became clear, a plan was formulated to organize the execution of the literature review. Okoli and Schabram (2011) discussed the plan as the protocol outlining the specific steps to conduct the literature review. The following section describes the protocol and its steps.

2.2. Implementation

Before conducting the literature review, it was important to define the protocol (Okoli & Schabram, 2011). The protocol of this study contains the following activities:

- Creating a "starting set" of articles
- Identifying the search concepts and terms
- Selecting the search engines
- Selecting the search strategy
- Analyzing and reviewing search results
- Creating a "final set"

Nine articles recommended by the research supervisors regarding the research theme "Data Science Goes Agile" were used as the starting set of articles to identify what concepts and terms can be incorporated into the search strategy. Table 1 below illustrates the key concepts, related terms, and combinations of these which are used to locate suitable articles.

Table 1: Search terms

	Key terms (AND)				
	data-driven	Life cycle	Agile	Methodologies	Project
Related terms (OR)	- Data science - Big data - Data mining - Knowledge discovery	- Process	- Software development - Agile principles - Agile practices - Agile implications	- Models - Methods - Frameworks	- Team - Stakeholders - Implications

The databases Open Universiteit (OU) Library and Google Scholar were used as search engines. The advanced search of the OU library made it possible to perform Boolean searches on the world's largest digital libraries in the field of IT, such as EBSCO Host. Google Scholar was employed for a wide-ranging search and to quickly view the related articles that could be noteworthy.

Searches began by running the queries in the search engines. The articles used for the literature review are scientific, peer-reviewed, limited by timeframe and contain specific concepts that are selected for the theoretical framework.

The PRISMA flowchart of Moher et al. (2009) was used as the selection strategy to reduce the number of search engine results until a final collection of articles was chosen as the final set. Three hundred thirty-eight articles were identified of which eight were duplicates, 305 were excluded based on the title and keywords, and another 12 were excluded based on their abstracts, introductions and conclusions. Thirteen articles were included in the final set. These articles were read thoroughly and were used as references for the literature review. The snowballing method was used for the articles in the final set that have relevant references but did not appear in the search results. This resulted in an expansion of the final set to 16 articles. Further information regarding the literature review are included in Appendix I- .

2.3. Results and conclusion

In this section, the results of the literature review are presented. The sub-sections relate to the research sub-questions and help formulate the answers to these questions.

Data development cycle

The focus of data-driven projects is on data, and the goal is to produce usable and valuable insights from the collected data, which are referred to as data products. Dawson et al. (2011) explain that data have a life cycle that begins at the moment the data are collected and ends at the moment the data become obsolete and lose their value. The cycle aims to establish the order of activities and processes involved in product development (Mariscal et al., 2010). Figure 4 illustrates several high-level activities within the data science development cycle.

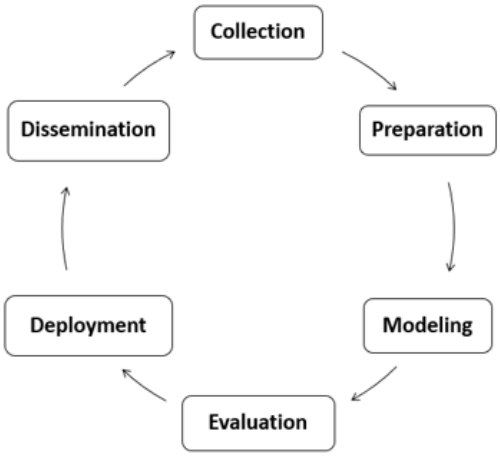


Figure 4: Data development cycle (Mariscal, 2010; Saltz et al. 2017)

Mariscal et al. (2010) reviewed the most used process models, in industrial and academic projects that are used in data and knowledge discovery projects. The result of the survey by Saltz et al. (2017) indicates that although these processes differ in detail, they are broadly similar at a high level. According to the most recent poll by KDnuggets.com (2014), CRISP-DM was the most widely used model, and since 2007, more companies used their own ad hoc process models for developing their data products. Marbán et al. (2008) discussed that even though CRISP-DM was considered as the "de facto standard" for developing data products, it was not as effective as it should be because it established a waterfall life cycle. Larson and Chang (2016) propose the "Agile Delivery Framework" in which the "fast analytics/data science" development cycle is used as a process model for developing data products. According to Larson and Chang (2016), the development cycle must involve iterative and incremental development to create, validate, and modify analytical models until the requirements are fulfilled.

Agile in data-driven projects

Agile is viable in and valuable to data-driven projects. Some studies argue that data-driven projects must focus more on people and how to more quickly and effectively operate the project than on process and technologies (Batra, 2018; Schmidt & Sun, 2018). Although developing data analytics applications, transforming data to actionable insights, and performing big data research are fundamentally different from Agile software development, there have been attempts to implement Agile in these projects (Jurney, 2017). The adaptive, dynamic, less formal, and customer-focused character of Agile appears to be an attractive answer to the challenges that DS projects encounter (Batra, 2018).

Values and principles

The Agile Manifesto is the summary of Agile principles. Table 2 emphasizes that while the four Agile values in the right-hand column are important, those in the left-hand column are even more important (Beck et al., 2014).

Table 2: Agile Manifesto (Beck et al. 2014)

More valuable statements		Less valuable statements
Individual and their interactions	Over	Processes and tools
Working software	Over	Comprehensive documentation
Customer satisfaction	Over	Contract negotiation
Responding to change	Over	Following the plan

Agile projects are people- oriented (Nascimento et al., 2012). The success of project is determined by the way in which individuals collaborate and interact, not the structured processes and tools (Larson & Chang, 2016). Unlike in traditional projects in which different disciplines work in different phases after one another, in Agile, all disciplines related to the project work together throughout the development cycle. Because of this, there is no necessity for comprehensive documentation (Larson & Chang, 2016). Through close collaboration among different stakeholders, Agile attempts to establish trust. The focus on close collaboration enables the exchange of knowledge and allows data-driven teams to explore possibilities for developing quicker and more qualitative data models and products (Larson & Chang, 2016). Agile welcomes and embraces changes throughout the project (Hoogendoorn, 2014). Iterations and small scale planning in Agile teams enable projects to respond to changes and not ignore them (Batra, 2018). Agile offers stakeholders the opportunity to modify requirements during the project instead of establishing them in advance (Larson & Chang, 2016).

The 12 principles of the Agile Manifesto are provided in Table 3; these principles present the characteristics of Agile projects and define the four values of the manifesto (Beck et al., 2001). As the table reveals, the top priorities of Agile projects include 1) satisfying the customer continuously, 2) embracing changing requirements, 3) delivering working software frequently, 4) having daily collaboration between the team and stakeholders, 5) motivating individuals and trust of team, 6) communicating face-to-face, 7) continuing to deliver a working product, 8) maintaining development pace, 9) focusing on quality, 10) creating simplicity, 11) self-organizing teams and 12) exhibiting continuous improvement.

Table 3: Principles behind of the Agile Manifesto (Beck et al., 2001)

#	Agile Principles
1	Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.
2	Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.
3	Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.
4	Business people and developers must work together daily throughout the project.
5	Build projects around motivated individuals. Give them the environment and support they need and trust them to get the job done.
6	The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
7	Working software is the primary measure of progress.
8	Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
9	Continuous attention to technical excellence and good design enhances agility.
10	Simplicity, "the art of maximizing the amount of work not done "--is essential.
11	The best architectures, requirements, and designs emerge from self-organizing teams.
12	At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behaviour accordingly.

Larson and Chang (2016) argue that Agile principles could address many of the common problems in DS projects. Batra (2018) and Larson and Chang (2016) illustrate that projects are embracing “close collaboration between project participants,” “active stakeholder participation,” “working on iterations” and “continuous delivery instead of comprehensive documentation”.

Practices

Practices are defined by TechTarget (2017) as “techniques that through experiences have proven to lead to the desired result reliably.” One of the impediments of Agile in data-driven projects is the adoption of the appropriate tools and techniques that would be of value for teams and enable professionals to work effectively. Various studies – such as Ambler (2012), Nascimento et al. (2012), Schmidt & Sun (2018) – have outlined best practices, listed below, which can be applied in a wide variety of situations and help data-driven teams work together effectively.

- Continued active participation and collaboration of stakeholders
- Just in time documentation, test-driven development, and Agile modeling
- Pair programming
- Continuous testing, deployment, integration, architecting, and improvements

Methodologies

Sekgweleo (2015) reported that Agile methodologies are based on iterative and incremental development and serve as a framework. The framework is a set of practices that assist the teams in developing a product that satisfies stakeholders’ requirements and needs. Agile methodologies share much of the same characteristics and practices; however, each has its own process, terminologies, and tactics. Agile has been used as an umbrella term that encompasses several methodologies such as Kanban, Extreme Programming (XP), Scrum, Spotify and Large-Scale Scrum (LeSS).

Although Agile is suitable in data-driven projects (Larson & Chang, 2016), the results of case studies by Batra (2018) and Saltz et al. (2018) indicate that the aforementioned Agile methodologies are not widely employed in these projects. Most data-driven organizations use their company-owned or hybrid methodologies for developing data products. Among those that do attempt to use Agile, Kanban is the most chosen Agile methodology, followed by Scrum. Batra (2018) states that, this does not mean that Agile is not used in data-driven projects; many data-driven organizations rely on customized approaches based on beneficial aspects of Agile methodologies instead of using the discussed methodologies.

Summary

The results of various studies indicate that linear and traditional approaches are the reasons data-driven projects are still failing. Schmidt and Sun (2018) revealed that most organizations choose to customize known and trusted data process models, such as KDD or CRISP-DM. The description of these approaches provides no guidance regarding how data-driven projects should function to increase the speed and quality of product development (Saltz, Shamshurin, & Croswston, 2017). Although Larson and Chang (2016) and Nascimento et al. (2012) prove that Agile ideas are suitable for data-driven projects, there are many obstacles that prevent organizations from adopting Agile. Larson and Chang (2016) and Schmidt and Sun (2018) note that available Agile data-driven frameworks do not include guides on the selection and adoption of Agile for a team. Most data-driven organizations combine traditional models with the useful aspects of Agile and develop their company-owned methodologies to operate their data-driven projects in an Agile manner. Table 4 provides an overview of how Agile is applied in data-driven projects.

Table 4: Agile applications in prior research

		HOW AGILE APPLIED			
		Process Development Lifecycle	Values and Principles	Practices	Methodologies & Frameworks
PRIOR RESEARCHES	(Nascimento et al., 2012)	- "Agile KDD lifecycle" (Iterative process)			"Agile KDD disciplines" framework
	(Schmidt & Sun, 2018)	- Knowledge Discovery (KD) steps/process (applied Intermingled and iterative)		-Incremental testing - Continuous integration - Test Driven Development (TDD) - Pair programming - User story - Scrum events (Applied throughout the KD steps)	"Agile-facilitated KD" framework (Interaction & Iteration)
	(Batra, 2018)	- More iterative and less structured product development process	- Shared understanding by communication - Technological capability (team capability and experience) - Top management commitment, leadership and governance - Self-organizing teams		
	(Larson & Chang, 2016)	Synthesizing "Fast Analytics/Data Science" lifecycle and Agile. (by doing iterative and incremental cycle through the following steps: Scope, Discover, Analyse, Model/Design/Develop, Validate and Deploy)	- Stakeholder collaboration (collaboration between product team, domain experts, business and technical experts)	- Small time-boxed increments - Co-located resources	"Agile Delivery" framework
	(Saltz et al., 2017)				- Applied "Kanban" and "Scrum" in data science projects

The adaption of Agile principles and the selection of the most valuable methodologies and practices for teams are significant complications of employing Agile in data-driven projects. Agile provides a manner of working. Applying Agile in a data-driven project requires more than a shift in the development process; applying Agile is related to the improvement of the development process, rapid delivery, increased and more effective communication, and higher quality. Traditional organizations with a corporate culture based on micromanagement have many concerns regarding the switch to the dynamic environment of Agile. Krawatzeck, Dinter, and Thi (2015) and Nascimento et al. (2012) indicate that there are many aspects to Agile in data-driven projects. It begins with project participants understanding the Agile Manifesto and principles and extends to using methods (e.g., techniques, tools) that support data-driven projects in the Agile way of working.

Conclusion

Agile requires a shift in mindset, a new manner of thinking, by every participant in project teams. All disciplines in a project must understand the mindset that is characterized by following the Agile Manifesto and its principles. This literature review found that the four values of the Manifesto (Table 2) and the 12 principles of Agile (Table 3) must be completely understood by project participants and must be adopted in every data-driven organization that attempts to implement Agile. In addition, an iterative and incremental development process together with Agile practices can support data-driven projects to improve the performance of their teams. Figure 5 illustrates the integration of the research questions and, themes that are inspired by the literature review; the theoretical framework on the right side answers the sub-questions of the research.

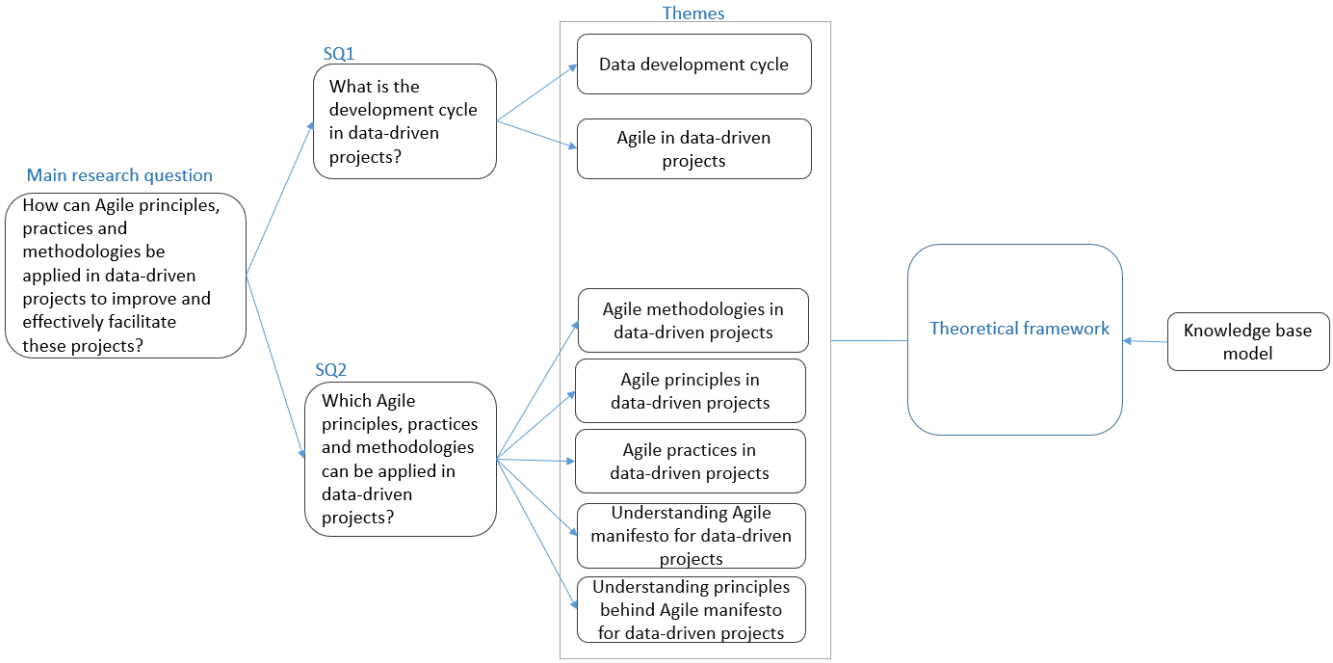


Figure 5: Theoretical overview

Figure 6 represents the theoretical framework that supports the DSR and provides the basis for future research. The “design and development” step of DSRM utilizes this theoretical framework to create a solution to the main research question and design an artifact that can lead data-driven projects to apply Agile. The Agile values and principles, the Agile methodology, and the set of Agile practices form the three pillars of the framework.

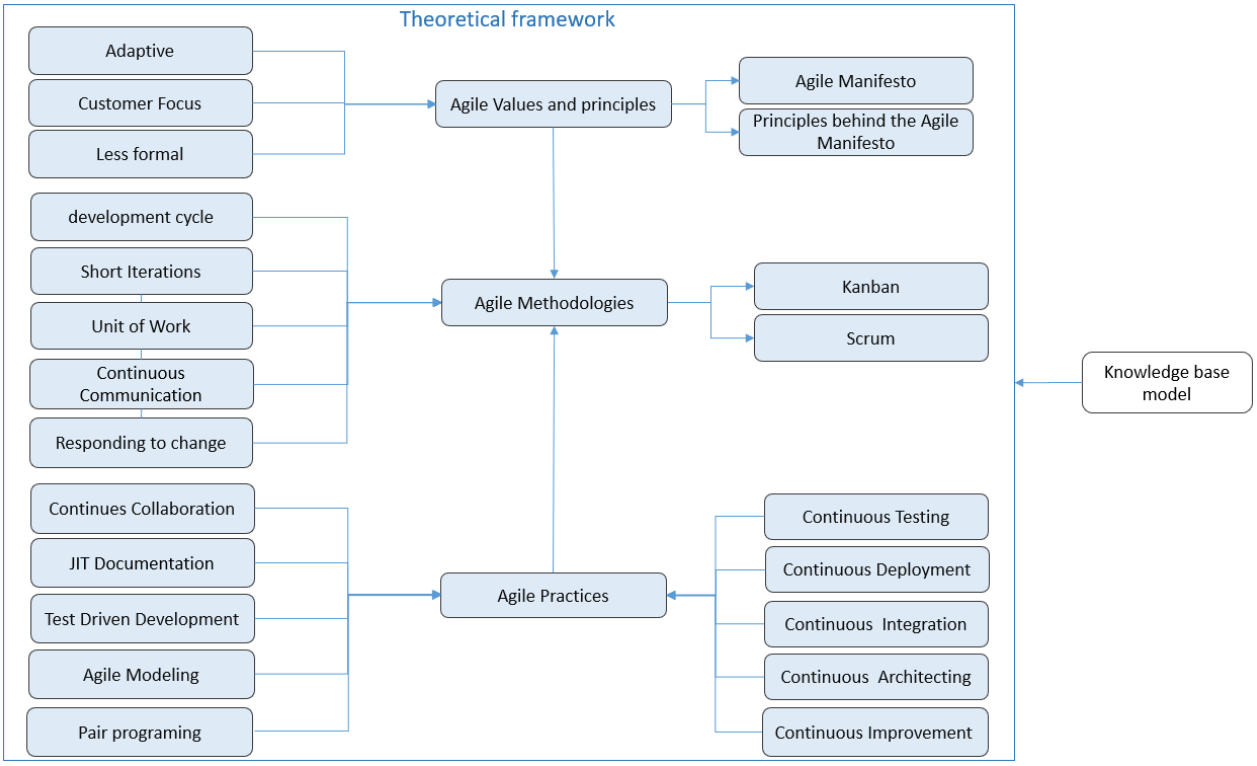


Figure 6: Theoretical framework

2.4. Objectives of follow-up

The lack of practical experience and maturity regarding Agile principles, practices, and methodologies in projects deters data-driven organizations from trusting an Agile way of working and applying it to their projects. Data-driven organizations require guidelines to determine which Agile practices and methodologies are appropriate and how to implement Agile to improve and effectively facilitate their projects.

Therefore, this DSR aims to 1) construct a set of guidelines and a framework that helps data-driven projects implement Agile 2) help teams select the applicable Agile principles and practices for their projects and 3) help teams apply these principles, practices to improve their development processes and performance.

3. Methodology

This chapter describes the DSR that is employed as the research approach. The reason for choosing DSR and the DSRM introduced by Peffers et al. (2007) is to create a proper design and solution.

3.1. Conceptual design

Saunders et al. (2012) mention that the selection of a research methodology depends on the philosophical stance of the research. Traditional research methodologies, which are informed by one or more of the five major philosophies discussed by Saunders et al. (2012), focus on confirming hypotheses by using data to test theories (quantitative approaches) or exploring the meanings expressed regarding a social or human phenomenon (qualitative approaches). In contrast to these traditional approaches, design science focuses on building and evaluating information technology (IT) artifacts to solve a problem in the field of information science (IS) (Hevner et al., 2004). This thesis employs DSR as the research approach since the main research question of this thesis is an open inquiry concerning an organizational problem in the field of IS and the goal of the research is to build an artifact as the solution to the business problem.

Kuechler et al. (2004) describe design science as a body of knowledge (in the form of well-developed theory, technique, method, and model) for developing applicable solution (artifact) to problems that satisfy specific desired goals. DSR is a set of activities that contribute to the design and development of relevant solutions in the field of IS (Peffers et al., 2007). Hevner et al. (2004) provide a DSR framework that illustrates seven guidelines used for understanding, executing, and evaluating DSR in IS. Figure 7 displays how the DSR framework is adopted for this research.

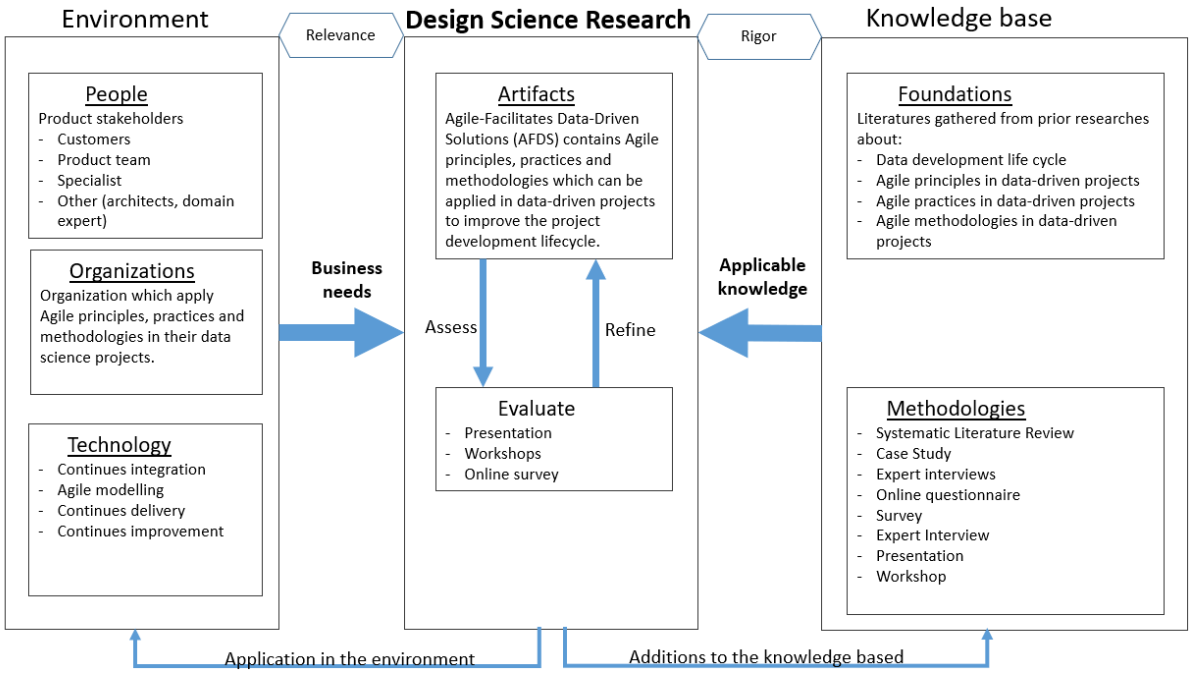


Figure 7: Design science research framework (Hevner et al., 2004)

Through the literature review, a "knowledge base" provides the theory that is used for the design and development of a solution in the form of artifacts. This study uses SLR, a questionnaire, expert interviews, and a workshop to evaluate and justify the scientific value of the design and validate the rigor of DSR. The "environment" is composed of people, organizations, and technology and defines the space wherein the research problem is formulated and the solution are utilized (Hevner et al., 2004).

The following section describes the DSRM and the steps taken to design and develop the artifact. These steps are drawn from the theoretical framework and provide answers to the main question of this thesis, which is discussed in Chapter 1.3.

3.2.Technical design

Peffers et al. (2007) introduced DSRM as an IS research methodology, this methodology is a process that incorporates principles, practices, and procedures needed to execute DSR. This methodology also enables researchers to provide an iteratively and incrementally applicable solutions to research problem (Peffers et al., 2007). The six activities of the DSRM process are illustrated in Figure 8, and the following subsections explain in greater details how they applied to this research.

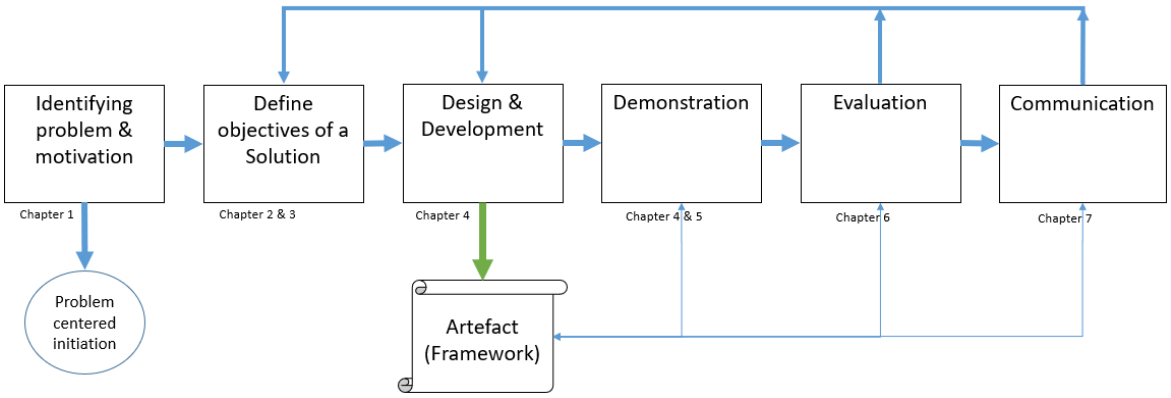


Figure 8: DSRM process model (Peffers et al., 2007)

Problem-centered approach

The result of various studies indicates that linear and traditional approaches are the reasons why data-driven projects are still failing. The description of these approaches provide no guidance regarding how data-driven projects should function to develop a product more quickly and with higher quality (Saltz, Shamshurin, & Croswston, 2017).

Problem identification and motivation

The previously conducted research has revealed that Agile data-driven projects can construct predictive models that can be continuously refined and improved (Schmidt & Sun, 2018). However, it appears that these projects have not yet gained enough experience with Agile and do not understand how to adapt it to their situations. Lack of knowledge about and experience with Agile causes the data-driven organizations to encounter daunting challenges in applying the Agile practices and methodologies that facilitate their projects most effectively.

Objectives of the solution

The objective of the solution is inferred from the problem statement and the theoretical framework. It indicates the possibilities and feasibility of the research. The research started with the literature review, which contains information from prior scientific research. The relevant literature was analyzed, and the outcome resulted in the theoretical framework depicted in Figure 6. The theoretical framework answers the sub-questions of the research and later in the research process, this framework provides the requirements for designing and developing the solution to the main research question.
The objective of this DSR is to design and develop an Agile framework as the design artifact of this research that supports organizations in implementing Agile in their data-driven projects. The framework consists of a set of guidelines that enables organizations to handle the challenges that they face in applying Agile in their data-driven projects. Design and development

This step aims to design the solution that answers the main question of this thesis. The design focuses on creating the artifact for solving a problem (Hevner et al., 2004). Kuechler et al. (2004) note that "an artifact can be any designed object in which a research contribution is embedded in the design." The artifact of this research is an Agile framework that can be applied in data-driven projects to improve the development process and the performance of their teams.
The theoretical framework from the literature review is utilized in this step to validate the requirements of the solution. The requirements are validated by conducting a questionnaire and five semi-structured interviews at a data-driven organization in the Netherlands. The company's primary business is management consulting in the field of strategy, structure, culture, resources, business processes and delivery solutions in the field of business intelligence (BI), big data, data science, data mining, and data warehousing. The questionnaire and expert interviews were employed to validate the requirements in the business environment upon which the evaluation of the framework was based.
The outcome of previous activities led to the design and development of the framework. During the presentation, the draft design of the framework was presented to the experts who were interviewed. When new requirements were identified, the design was modified to enhance the framework. In chapter 4 the design and development of the framework are further described.

Demonstration

The demonstration is the application of the solution to the problem domain (Peffers et al., 2007). Due to time constraints, the demonstration of the framework was only consisted of using a presentation. After the presentation, the framework was applied to solve a business problem by conducting a hands-on workshop with the consultants of the company being researched. The attendees participated in the workshop, where they worked on solving a fictional use case in order to practice applying the AFDS framework. The goal of the workshop was to determine a solution, which is presented in Figure 16: AFDS Development method, to a business problem by using the framework. The results of the demonstration are described in chapter 5.

Evaluation

The evaluation strategy selected for this thesis is based on the framework for evaluation in design science (FEDS). Venable et al. (2016) introduced the FEDS alongside a process to guide researchers in evaluating the artifacts that were designed during DSR projects. The evaluation strategy selected for this thesis contributes to both dimensions of the graph depicted in Figure 9.

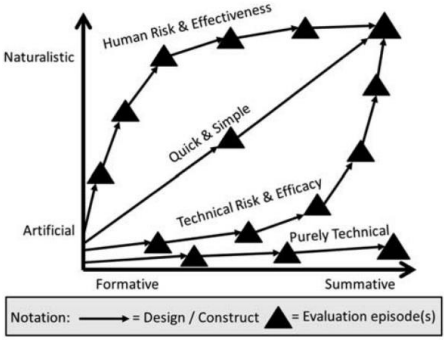


Figure 9: FEDS with evaluation strategies (Venable et al., 2016)

A naturalistic evaluation that explores the performance of the design artifact within an organization is represented as a case study (Venable et al., 2016).

A formative evaluation aims to help improve the outcome of a design process. The formative evaluation of this thesis was conducted using an online questionnaire, expert interviews, and a

validation presentation during the design and development stage. Further details about the formative evaluation and the results are provided in section 6.1.

A summative evaluation, in conjunction with a formative evaluation, enables researchers to test frameworks and ensures the reduction of risks to the users of the framework (ethics) and acknowledge the results. The summative evaluation of the framework of this study is conducted by demonstrating the framework at a presentation followed by a hands-on workshop. The results of the summative evaluation are further described in section 6.2.

Prat et al. (2014) mention in their study that design science researchers often struggle with how to evaluate the IS artifact. The study introduces a holistic view of evaluation criteria for a design science artifact; this view is depicted in Figure 10, which also provides a list of criteria for evaluating the IS artifact. Prat et al. (2014) organized the evaluation criteria along the five fundamental dimensions of a system, which are here considered as the design science artifact: goal, environment, structure, activity, and evolution.

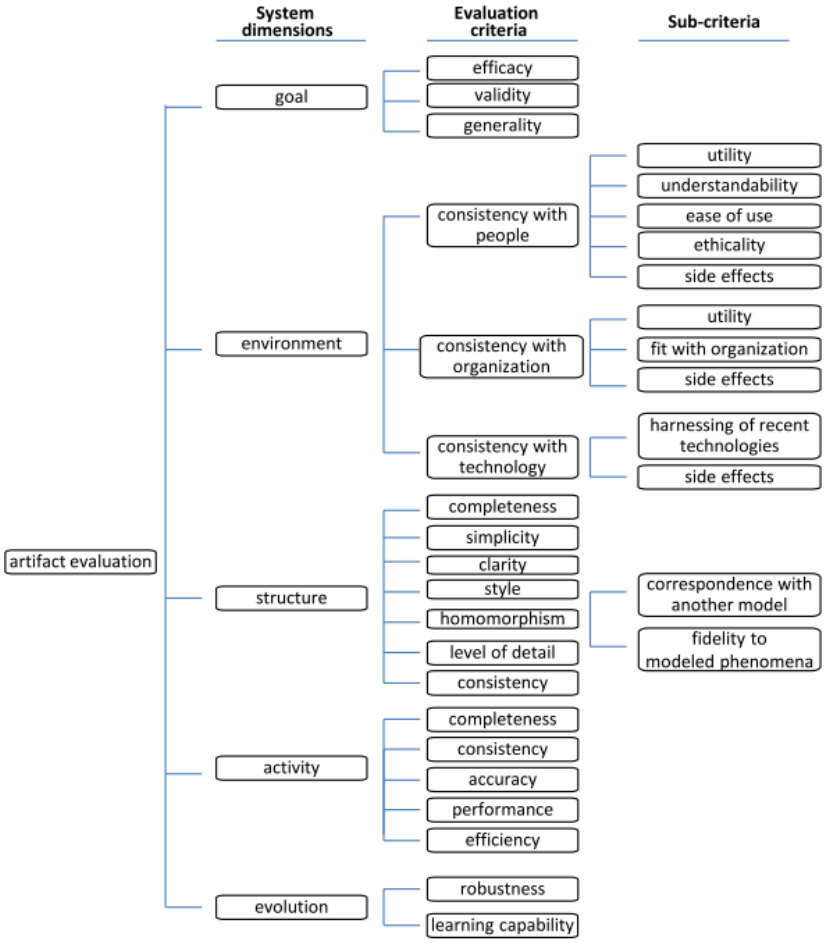


Figure 10: Hierarchy of criteria for IS artifact evaluation (Prat et al., 2014)

The details of the evaluation of the framework are described in chapter 6.

Communication

The final step of research is communicating the problem, the solution, and the solution's design (Peffers et al., 2007). The research of this thesis concluded with the communication of the designed framework through a, thesis presentation, and the publication of this article. Figure 11 illustrates how the DSR is applied in this thesis:

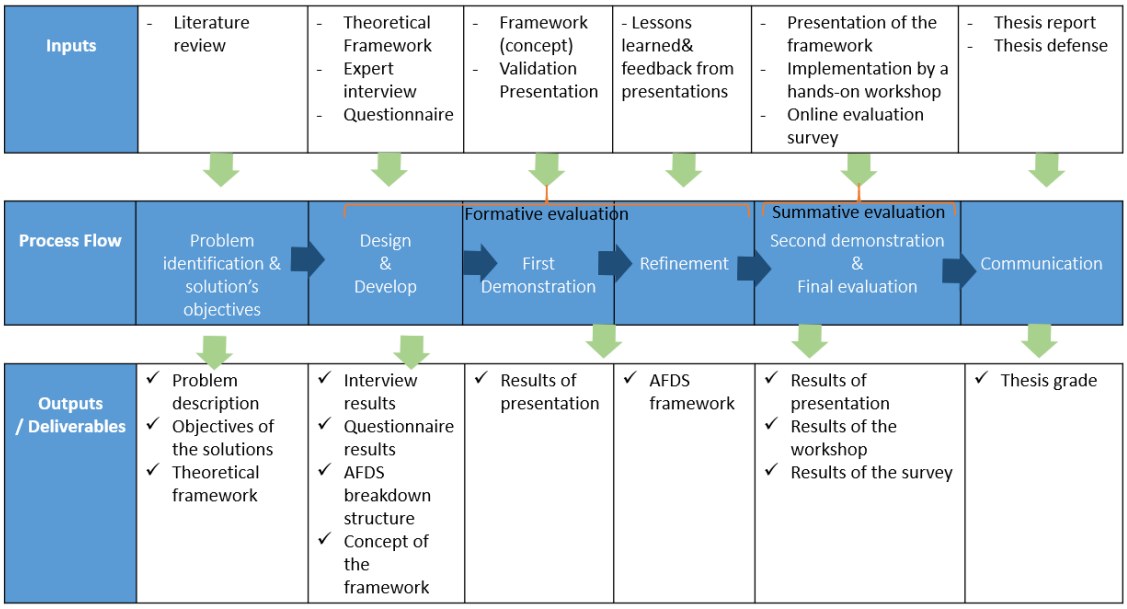


Figure 11: Research process and deliverables

3.3. Reflection relevance and rigor

While traditional research approaches focus on the reliability and validity of confirmed hypotheses, DSR focuses on rigorously evaluating the utility, quality, and efficacy of design artifacts via well-executed evaluation methods. Hevner et al. (2004) describe DSR as an iterative and incremental problem-solving process that aims to connect rigor (scientific domain) and relevance (business domain). The DSR of this thesis aims to create an artifact to solve real-life problem in the field of IS.

The rigor cycle provides prior knowledge to the DSR project to ensure that the designed artifacts are scientific contributions. The relevance cycle provides the requirements from a business environment and defines the acceptance criteria for the ultimate evaluation of the research results. The design cycle is recognized as the core of any DSR, and it is iterated between the design and development of the artifact and its evaluation. Figure 12: Design Science research cycles *(Hevner, 2007)*Figure 12 offers a viewpoint of the DSR framework (which is illustrated in Figure 7) with a focus on three inherent research cycles represented by Hevner (2007).

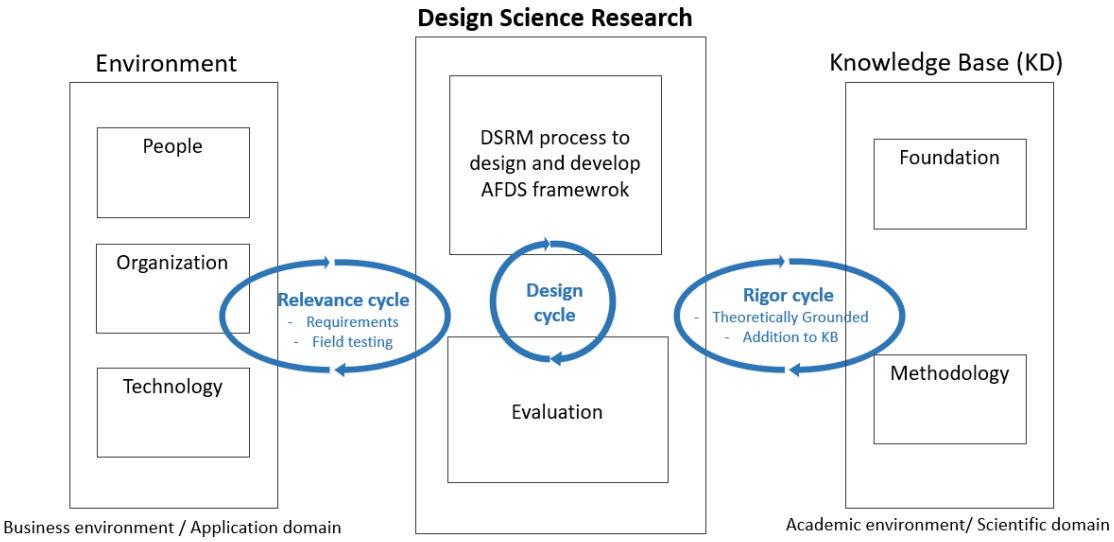


Figure 12: Design Science research cycles (Hevner, 2007)

In order to ensure the rigor of the designed artifact (rigor cycle), the DSR project of this thesis uses the theoretical framework from the literature review and SLR, a questionnaire and expert interviews. The experts, the company's clients, and technologies that are used in the projects define the business environments where the research problem was formulated, and the results were evaluated (relevance cycle). The questionnaire, expert interviews, presentations, and the hands-on workshop are the techniques that were used for the formative and summative evaluation of the framework.

Expert interviews and a questionnaire are used to validate the requirements in the business environment and evaluate the correctness of the framework during design and development (formative evaluation). The experts were consultants of the data-driven company where the research project is executed. They consisted of a 1) principle manager who is one of the two directors of the company, 2) principle expert who is the other director, 3) data solution architect who provides consultation with his client in the transport sector 4) a scrum master/BI consultant who works in a scrum team and provides consultations with his client in the health care sector, and 5) a senior tester who works in a scrum team and assists his client in the finance sector by applying Agile test techniques and practices.

An interview protocol is used in the form of an interview guide during the expert interviews. This can be found in Appendix II- Interview protocol.

The questionnaire was sent out to all of the company's consultants (20 individuals, of which 14 participated). The questionnaire can be found in Appendix III- Questionnaire. The results of the formative evaluation are included in section 6.1.

When the design was completed, the framework was demonstrated through a presentation to the company, followed by a workshop. Details about the demonstration are included in chapter 5 and the results are in Appendix VII- Results of the evaluation survey.

The DSR study is complete and effective when it satisfies the requirements and constraints of the problem it was intended to solve. The results of the evaluation of the framework are included in chapter 6.

4. Design and development

This chapter focuses on designing and developing the “Agile Facilitates Data-Driven Solutions” (AFDS) framework. The AFDS framework aims to change the linear and plan-driven characteristics of data-driven models, such as CRISP-DM and facilitate data-driven organizations with Agile principles, practices, and methodologies to improve the development process and maximize productivity when developing their data products. Figure 13 illustrates the mapping between the AFDS framework and processes mentioned earlier.

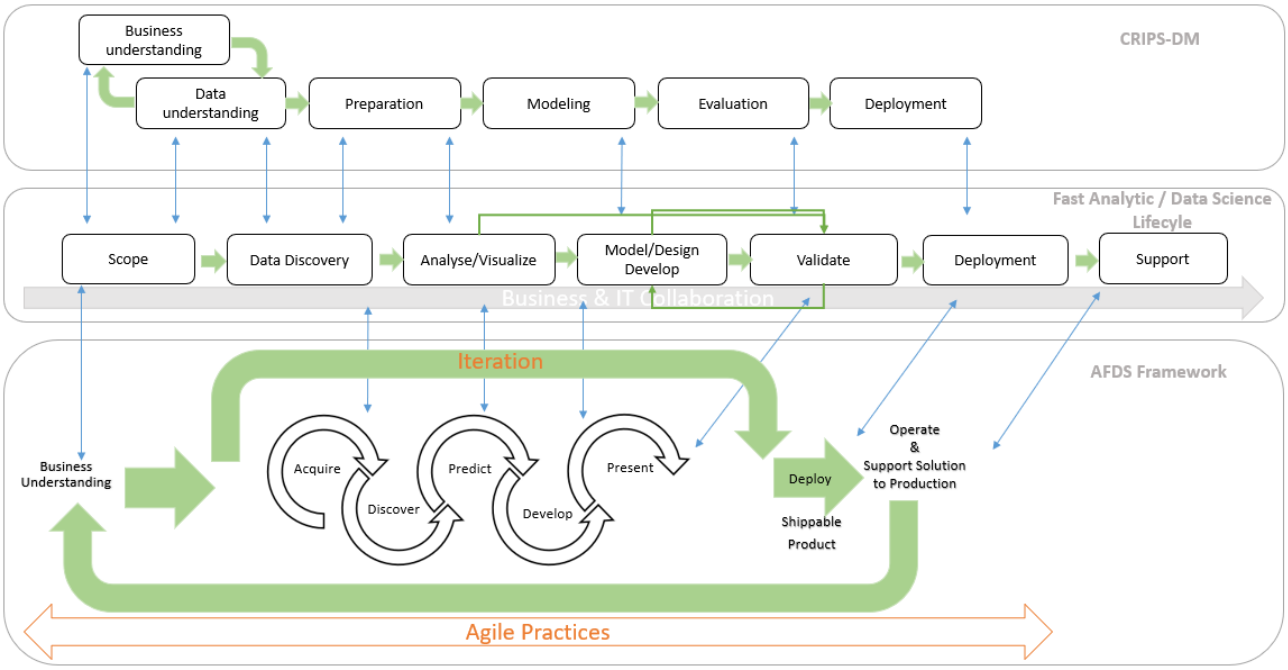


Figure 13: Mapping AFDS Framework with CRIPS-DM & Fast Analytics/Data Science

4.1. Agile Facilitates Data-Driven Solutions (AFDS) framework

Agile challenges the traditional, sequential and linear data process approaches. This is expressed in various ways in Agile’s values and principles and is comprehensively addressed in the literature review. Although there are multiple lifecycle steps, they should not be taken in a certain fixed order. There should be Agility to go back and forth one or more steps as often as needed. Agile data-driven teams should be able to alternate between the steps as needed. Instead of having the focus on following a plan, the focus is on having a working product. Instead of waiting until all requirements are implemented before the stakeholders are presented with a solution, the stakeholders should be involved every step of the way.

An iterative approach suits Agile’s values and principles. Iterative approaches motivate teams to work continues, at regular intervals, at a constant pace and frequent within a timescale. This is reflected in the description of the principles. Principle #1 speaks about “early and continuous delivery.” Principle #3 refers to ”frequent delivery… with a preference [for] the shorter timescale.” Principle #8 states that “sponsors, developers and users should be able to maintain a constant pace.” Principle #9 discusses “continuous attention,” and Principle #12 is about reflection at regular intervals. An iteration is defined as the repetition of a process applied to the result of a previous application as a means of sequentially approaching to the solution of a problem.

Working in iterations embraces the Agile way of working and addresses the problems data-driven projects experience with linear approaches.

The AFDS framework targets the entire project organization and helps teams create small shippable data products in short iterations. This framework also helps them inspect how the products are created. This contributes to the first objective of this thesis. Figure 14 illustrates the framework breakdown structure. The framework adopts some of the Scrum guidelines to provide an iterative and incremental method and to develop a data product in short iterations. The framework also suggests a development cycle that is extracted from the fast analytics/data science lifecycle (Larson & Chang, 2016), and proposes a set of Agile practices, that can be used by multidisciplinary teams. This contributes to the second objective of this research.

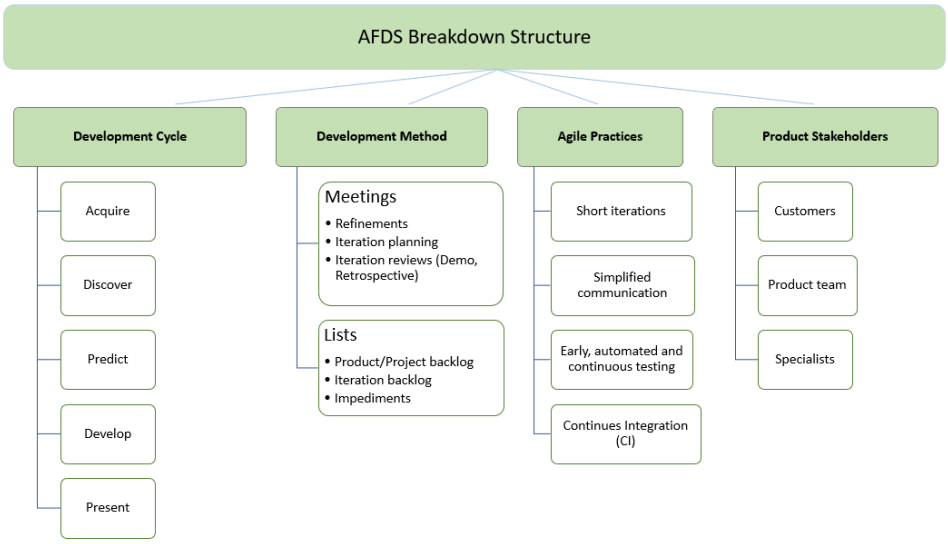


Figure 14: AFDS framework breakdown structure

AFDS data development cycle

The aim of the AFDS development cycle, as a part of the AFDS method, is to create a clear understanding of the development cycle of a data product in an Agile project environment and support multidisciplinary teams in creating actionable insights and developing data products. Figure 15 illustrates AFDS development lifecycle which is extracted from the CRISP-DM process model and fast analytics/data science lifecycle (Larson & Chang, 2016).

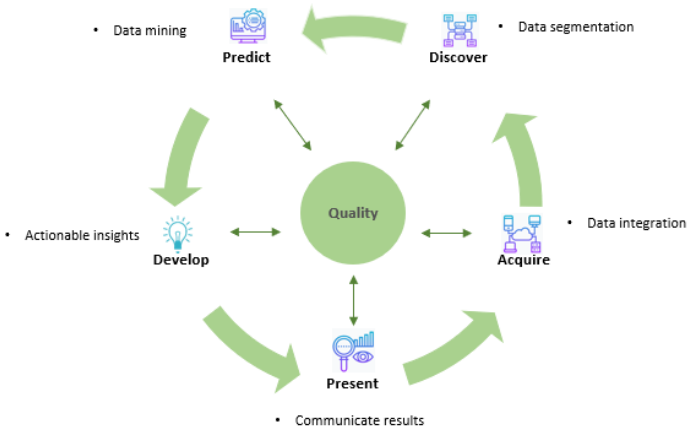


Figure 15: AFDS data development cycle

Acquire
In this stage, the data from various resources are gathered, boundaries are defined, and expectations are recorded. Using solutions to store unstructured data creates more availability, flexibility, and quality. Storing data from internal and external sources enables access to unstructured (raw) data at an early stage of the process (Availability). Various disciplines in a product team such as a data scientist and a data analyst may access the data any time they need (flexibility). They may trust data when they get it directly from the data lake (Quality).

Discover
When data is acquired, the next step is to discover the value of data, highlight the relationships between variables and identify elements which can be used in analytical models. Technologies such as data segmentation enable to identify and categorize data in specific sections to use data more efficiently.

Predict
In this step, the data turns into useful information. In analytical modelling, machine learning algorithms such as regressions, clustering and classifications are used to make descriptive, predictive, and prescriptive analysis models. Modeling in this step focuses on pattern recognition of the data as well as analytical modeling (Larson & Chang, 2016).

Develop
Developing and testing of the data product is encapsulated in the one step. The focus is on producing a product that creates actionable insights. Testing is about validating the product to ensure it meets the requirements, responds correctly to all kind of inputs and achieves the result which the stakeholders desire. Automated testing, interface black-box testing, database white-box testing, data quality testing, A/B testing, regression testing and performance testing are techniques that enable the teams to test their product and ensure accurate and reliable insights are drawn.

Present
When the product passes the tests and is ready to be delivered to its user's, various user interface applications such as dashboards are used for communicating and presenting the results to the stakeholders.

Quality
Integrating the quality to the development cycle and scale it across the process creates trust and quality management throughout the development cycle. Tools such as Collibra provides quality management across the development cycle and enables the DS teams to quickly and securely deliver trusted data products to who may need it.

AFDS development method

The results of case studies by Batra (2018) and Saltz et al. (2018) indicate that Scrum is not widely employed in data-driven projects. Data products are far less definable and therefore it is more difficult to plan the development cycle. The preferable results are not always clear at the beginning.

The progressive insights during development determine the result and actionable insights. The faster this progressive insight are discovered, the quicker the course can be altered to optimize the result. In order to provide iterative and incremental product development in short iterations, the AFDS framework incorporates the AFDS development cycle, is explained further, with Scrum as an Agile framework as presented below in Figure 16.

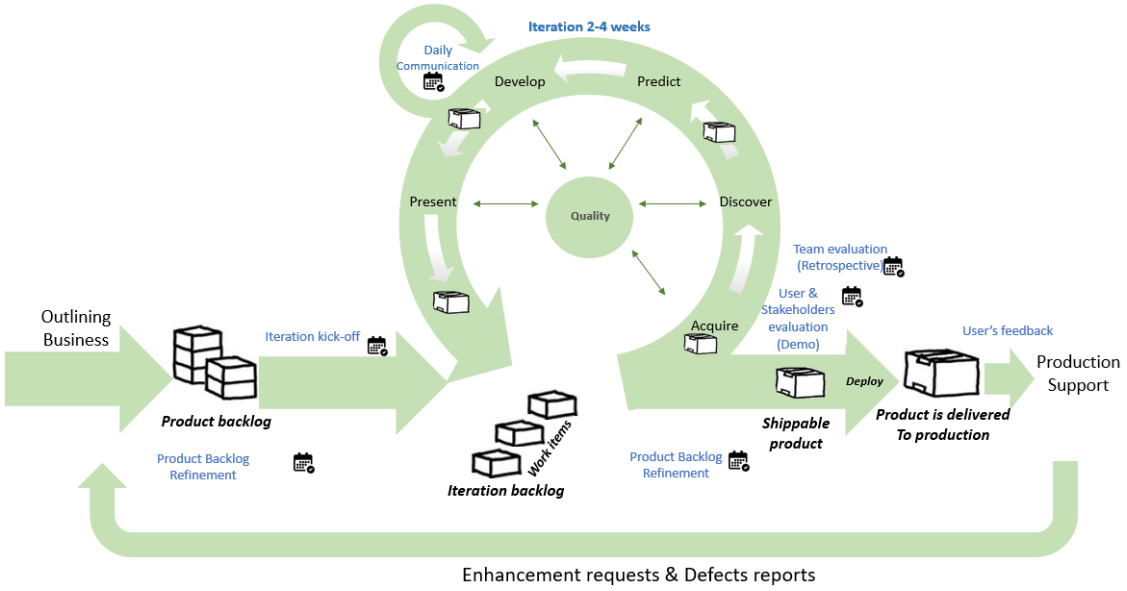


Figure 16: AFDS Development method

The process of AFDS method starts with the Agile practices 'Outlining Business' in which all stakeholder's requirements (functional, non-functional and constraints) are gathered, translated into the work items moved to the product roadmap and further into the Product Backlog.

Hoogendoorn (2014) mentioned, "The product backlog administers all the work in the project". The items in the product backlog are usually not detailed. The Product Owner (PO) fills the product backlog with the work items and discusses each item with the team during the backlog refinement to ensure that enough knowledge is available for the realization.

During the iteration planning, the work items that are selected for the next iteration and are ready to be implemented, are moved on to the iteration backlog. Each work item will go through all the activities in the development cycle: acquire, discover, predict, develop, and present. Once the work item satisfies the conditions based on the Definition of Done (DOD) and the PO has accepted it, the realization is done. At the end of each iteration, the realized work items are deployed to the staging environment to get final approval from the stakeholders. Once all the stakeholders are satisfied with the solution, the operations team releases the solution into production.

The support team is responsible for providing enhancements and bug reports to the development team. During the evaluation meetings, the stakeholders express their satisfaction or dissatisfactions about the product and the team about the development process. The feedback is noted and used as an improvement for next iterations.

The foundation of AFDS method is established on the Manifesto of Agile software development (Fowler, 2004) and Agile principles (Beck et al., 2001) and relies on the three pillars of Scrum (Schwaber & Sutherland, 2011) that are: transparency, inspection, and adaptation. It is significant that the work progress is visible for everybody who's working on the project and everyone is aware of the ins and outs (transparency). Product users and stakeholders are involved during the development cycle and are aware of the work progress. Their presence and feedback at the end of each iteration are significant for the development team to deliver a high-quality and useful product (inspection). When the work has no progress, or the result is not as desired and expected by its stakeholders, the adjustment of the process or the product will be made as soon as possible to minimize deviations (adaption).

AFDS Agile practices

TechTarget (2017) defined best practices as "techniques that through experiences have proven to lead to the desired result reliably." Agile practices below can be applied in a wide variety of situations and help data-driven projects to apply Agile in their teams.

Short iteration

Iterations break the linear approach. They provide the project with the chance to have early results and gather quick feedback. During each iteration, a part of the product is delivered. The duration of iterations may range from one to three weeks. Iterations of two weeks are strongly recommended, especially for beginner Agile teams. Each iteration is composed of three parts; it starts with an iteration kick-off to determine which work items will be realized during the iteration (e.g., filling iteration backlog). The main goal of an iteration is the realization of the agreed work items. At the end of an iteration, evaluation meetings take place. The purpose of these meetings is to evaluate the realized work items with stakeholders (Demo and Retrospective) (Hoogendoorn, 2014; Schwaber & Sutherland, 2011).

Simplified communication

In an Agile team, communication needs to be transparent and simple. The following practices and tools for this matter are:

Daily Communication

It is crucial that all the roles in a team collaborate closely with each other. Distributed Agile project teams, in which the team members are in different places, need to apply video conference tools such as Skype or Google Hangouts for daily communication, iteration planning, and evaluation meetings. Continuous communication is essential. Therefore, it is important to keep communication as simple as possible (Hoogendoorn, 2014).

Unit of work

The only way to deliver fully realized work items in each iteration is to use a unit of work, e.g. User story or Smart use case. A unit of work should be independent, negotiable, valuable, estimable, small, testable, realizable and acceptable (Cohn, 2004).

Visualize Iteration workflow

For a proper understanding of the amount of work in progress, the backlog needs to be visualized. Generally, each Agile project team has a board. Figure 17 represents the steps that are chosen by AFDS to visualize the flow of the work items during an iteration. Different teams chose different workflows. Many teams use "To Do, In Progress and Done" as a workflow. There are various tools available, e.g. Trello, that Agile teams can use to make a digital project board.

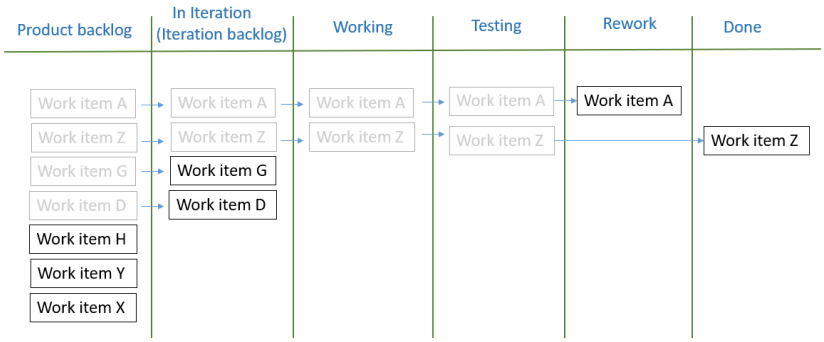


Figure 17: AFDS Iteration Workflow

The incomplete items in each iteration are considered as a new work item for the next iteration.

> ***Agile Documentation***
> Ambler (2012) introduces Agile Modeling practices and techniques for effective documentation in Agile projects. The techniques and practices such as TDD (Test Driven Design), requirements envisioning, architecture envisioning, documenting continuously, requirements prioritization and a single source of information is recommended by the AFDS framework to documents insights as it occurs during the development process.

Early, automated and continuous testing

In order to frequently deliver a data product and ship intermediate results, Agile teams need to test early and often. Automated testing enables teams to start testing early. Batarseh & Gonzalez (2018) describe various automating test techniques such as Analytics Driven Testing which support teams to ensure accurate and reliable products are created. Automated unit testing tools such as the Unit test library for Python developers (The Python Standard Library, 2016) enables teams to test the smallest unit of the product automated as early and as often as possible.

Continues Integration (CI)

Continues Integration (CI) and Continuous Deployment (CD) from Agile practices are techniques that support teams to iteratively publish their products and intermediate results.
Continues integration requires that all the code, data, and documents from different sources are integrated into a shared version control system multiple times a day. Build servers such as Travis CI, Circle CI or Jenkins enables the teams to build, test and deploy the product automatically as often as possible. Continues and iterative sharing of work, even when it may not be completed, enable teams to embrace and adopt changes and stay productive.

Continuous Improvement

Continues improvement, is one of the core Agile practices which is added to the AFDS framework. Based on the received feedback from the evaluation meetings, the team needs to immediately take actions to apply the feedback. This evaluation aims to learn from the lessons and to improve. Improvement can be related to the process, e.g. changing the duration of iteration, customizing the workflow board, or can be related to the product, e.g. changing the data privacy requirements, changing the business rules, or introducing and applying Agile best practices to the team such as Pair programming to improve the collaboration between the data scientist and software engineers and increase the quality of the data product.

AFDS method together with the ADFS development cycle and Agile practices contributes to the second objective of this research.

AFDS Product stakeholders

An Agile multidisciplinary team is a collection of individuals with their own expertise. It consists of all disciplines and roles needed to realize the work items in each iteration. What roles are needed in the team vary from project to project and mostly depends on the organization structure (Hoogendoorn, 2014). The AFDS framework defines the roles below in the project shown in Figure 18.

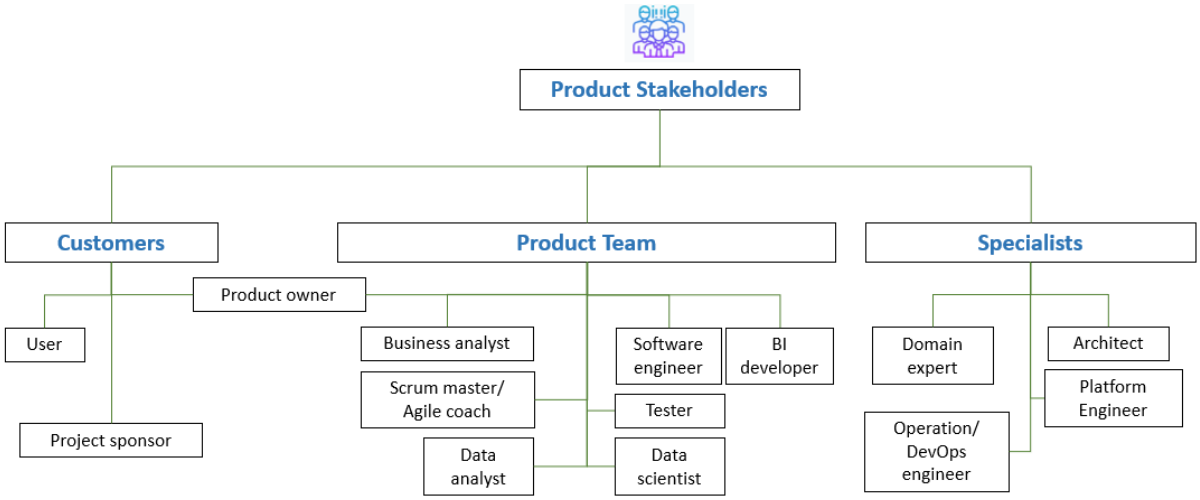


Figure 18: AFDS project stakeholders

All above Agile project roles on the data-driven project, which consists of the customers, the product team and the specialists need to collaborate closely with each other in a multidisciplinary project team, to deliver data products and actionable insights with business value in every iteration. A multidisciplinary team supports the Agile value of “individuals and interactions over processes and tools” and principle #4, which states that “business people and developers must work together daily throughout the project.” Working together on one team enables teams to deliver a fully functional product quickly; furthermore, problems can be detected early by working together.

5. Demonstration

The demonstration of the AFDS framework was a presentation followed by a workshop with the consultants of the company being researched. The two company directors, the BI consultant/scrum master and the tester were present at the presentation. There were four hours reserved for the demonstration session. The results of the demonstration are described below.

5.1. Results of the demonstration

Presentation

The AFDS framework, which is described in section 4.1, was presented to the attendees. Some discussion arose during the presentation, which is summarized in this section.

Development method

The AFDS development method, which is illustrated in Figure 16, is the answer to the discussion about how the process is running from the business question until the team built the solution incrementally and delivers it iteratively.

The director summarizes the process of the AFDS development method as follows:

> *What you are actually doing here is starting with listening and talking to the business to find out what the question of the business is, and the business wants. When the problem is clear and you have an idea about the solution, then you divide the solution into the number of work items for the product team to enable the interactive and incremental development process. When the product is demonstrated to its stakeholders and accepted by the business and the user, it is deployed in production. In support and operation, the user feedback and any possible production issues are analyzed, determined, and reassigned to the business.*

There was a short discussion about the operation and support activities as a part of the AFDS development process and comparing it with DevOps as an Agile framework. To keep the AFDS framework as simple as possible, DevOps was not used in this framework. Implementing DevOps as an Agile framework with all its technologies and building blocks is a major challenge for software development. DevOps is more than combining the development teams with the operation team. Considering AFDS is based on the Scrum framework and not DevOps, this discussion is kept outside the scope.

Development cycle

Director II indicated that he has doubts about the AFDS development cycle and whether it is suitable for data science product development. He believes that the development of a data product is certainly separated from what the data science does. He said that “it’s very important to define the process that fits within the work that needs to be done. I think this cycle is appropriate for big data, data analytics, and BI development but not for data science.” Director one said as a response to his colleague, “I think the potency of this cycle is bringing all disciplines together in a multidisciplinary team in which each discipline can do its work with others independent during this development cycle.”

Agile practices

There were no specific comments and feedback about Agile practices during the presentation.

Stakeholders

There was a discussion about the definition of the role of a data scientist. Director II indicated that "*it is very important to define this role and the tasks of a data scientist in the development process*." All of the experts agreed that the data scientist is a role that includes researching data and finding actionable insights. Where in AFDS development process the data scientist will be involved, is strongly depended on the business question and the problem that needs to be solved.

Hands-on workshop

After the presentation, the attendees participated in a hands-on workshop in which they worked on implementing a fictional use case for 45 minutes in order to practice implementing the AFDS framework. Each of the participants was assigned a position based on AFDS project stakeholders. The positions were defined as followed: 1) a product owner (PO), who represented the business; a scrum master (SCM), who supported the team during development cycle for adapting Agile; a data scientist; a BI consultant; and a tester. The goal of the workshop was to define and deliver a solution for a problem of the business by using the AFDS development method, which is presented in Figure 16: AFDS Development method. The business problem, which was presented by the product owner, was as follows:

"A general practice office has major complaints about long waiting queues."

In a 10-minute timebox session, the product team asked the product owner questions to understand the problem and create a solution together with the product owner. Once a solution direction was formed, a user story was defined by the product owner for the team as follows: "As a general practice, we want to see the waiting times of our doctors on a digital dashboard so that we get a better insight into the waiting time."

Once the user story was ready, the team started the realization by following the development cycle, which is illustrated in Figure 15: AFDS data development cycle. They were two 15-minute iterations defined to develop the solution. The team had to deliver a part of the solution in iteration 1 and the whole solution in iteration 2.

After this session, the framework was evaluated by the attendees with an online survey, which is included in Appendix VI- Evaluation survey. The four elements of the AFDS framework were evaluated based on the evaluation criteria, which are discussed in the next chapter. The results of the survey are included in Appendix VII- Results of the evaluation survey.

6. Evaluation

This chapter presents the results of the formative and summative evaluations of the AFDS framework. Figure 19 illustrates the structure of the evaluation of the AFDS framework.

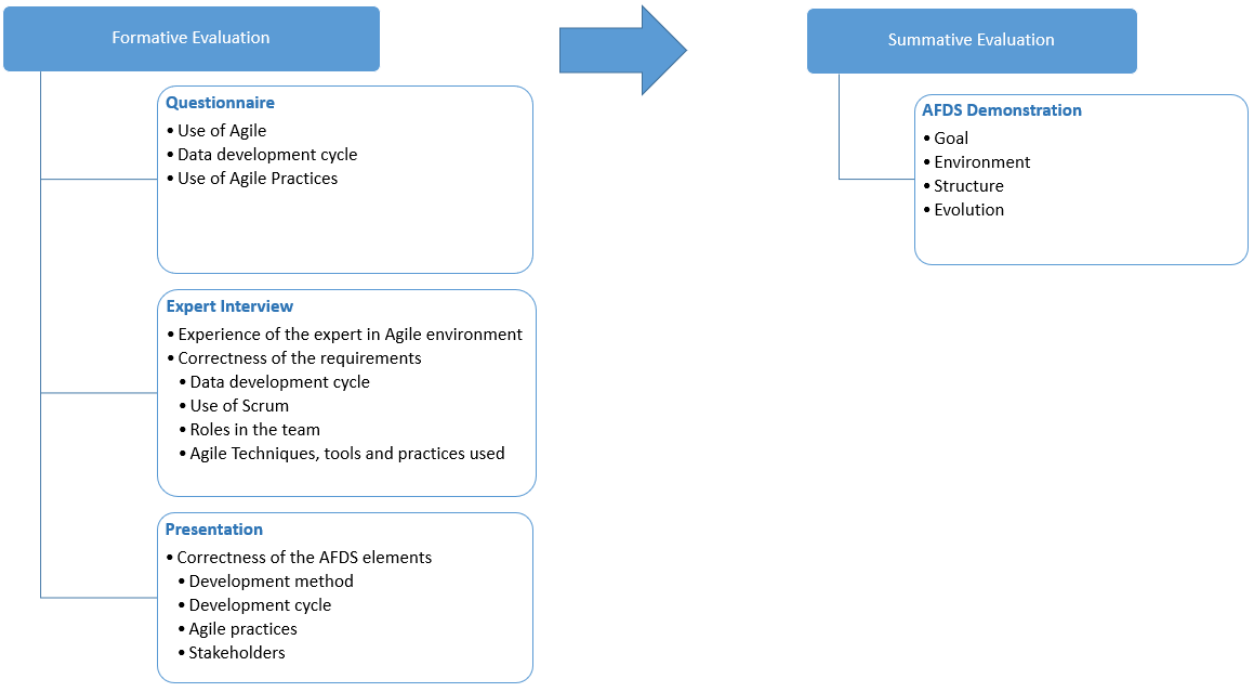


Figure 19: AFDS evaluation structure

Section 6.1 presents the results of formative evaluation and section 6.2 presents the results of the AFDS demonstration to the organization, which is considered as the summative evaluation strategy.

6.1. Formative evaluation

The questionnaire, expert interviews, and validation presentation were employed to validate the AFDS requirements in the business environment, upon which the formative evaluation of the framework was based.

The observation of the research problem in the business environment and validation of the requirements took place through the online questionnaire (Appendix III- Questionnaire) and expert interview (Appendix II- Interview protocol).

For the validation of the framework during the design and development, the concept of the designed framework is presented to the experts during a presentation. This presentation is used as the strategy to verify designed artifact and get earlier feedback for further improvement of the design.

In the next subsections, the results of the requirement validation and design verification are described.

Requirements validation

The aim of the requirements validation, during design and development, was to ensure that the requirements are clear and understood and meet the needs of business and achieve the business objectives.

Online questionnaire

The online questionnaire was sent out to 20 consultants in the company being studied. Of these 20 consultants, 14 participated. The results of the questionnaire are used for the observation of the problem in the business environment and the validation of the requirements during the design and development phase.

The questionnaire is constructed based on the literature review and uses the following evaluation criteria: the use of Agile, the data development process, and Agile best practices. The results of the questionnaire indicate that 85% of the consultants work in an Agile environment and use Scrum.

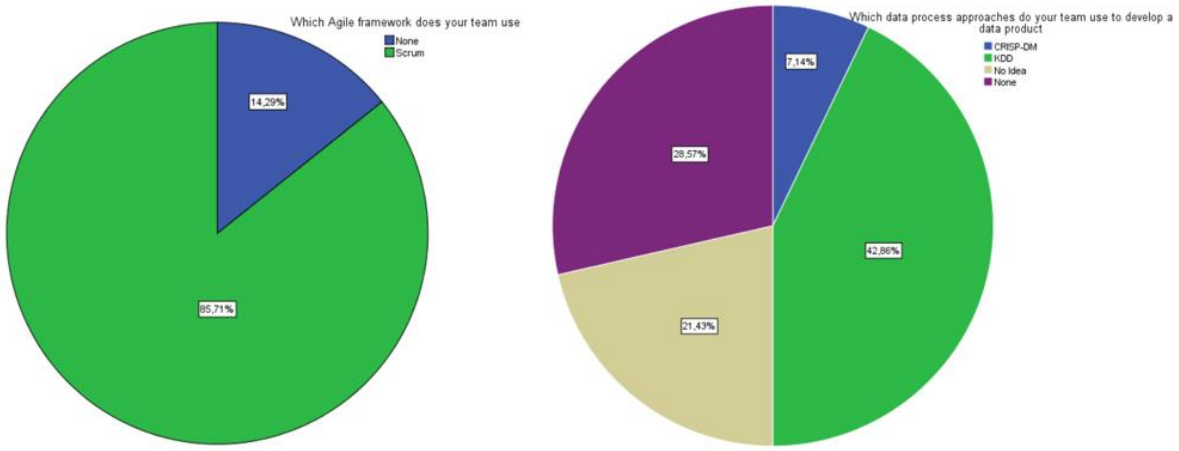


Figure 20: Use of Agile and use of data process approaches in business environments

Out of the 14 consultants, 42% stated that that they use KDD as the data development process in their project, 29% did not know the methodology used, and 21% indicated that they did not use a methodology. One consultant mentioned the Common Sense Development Program as the approach used for the development process.

The respondents indicated the best practices used as Agile in their projects are pair programming, continuous improvement, continuous active participation and collaboration of stakeholders, just in time documentation, automated testing, continuous integration, continuous deployment, test-driven development, Agile modeling, and continuous architecting. The results of the questionnaire indicate that the consultants considered continuous improvement as the most important best practices for data-driven projects.

Expert interviews

The expert interviews (Appendix V- Results of expert interviews are used as the formative evaluation strategy to validate the requirements of the framework during the design and development phase. The aim of this validation is to achieve a clear understanding of the research problem in the business environment and to validate the requirements to ensure that the framework meets the business objectives.
In this subsection, the evaluated themes are explained and illustrated with the quotes of experts from the transcripts.

Use of Agile principles

All interviews indicated that the Agile way of working is applied in their projects. One of the directors indicated that “Agile in data-driven environments is about dealing with constantly changing and uncertain circumstances, which are based on transparency, mutual trust, and close cooperation between all product stakeholders.” He said that the behavior of the team members and the

organizational principles, structure, and agreements are connected to each other and influence each other.

I think an Agile framework has a lot to do with people in a team and a team can be a project team, but that can also be an organization and that is very important if you want to change an organization. On the other hand, I also believe that the behavior of people in such a team, is very much determined by the organization. So, they may seem like a contradiction but if you look closely, they indicate each other (Director I/Principle manager).

The principal expert mentioned the literal meaning of Agile: "Agile means maneuverable." He also indicated that in data-driven projects, it is important to understand what the customer wants and how to achieve the desired results quickly without detailed up-front planning:

Always look at what goals we have and what the customer wants. I am in favor of not going to work with very big plans because you can always then change the plan. Certainly, if you are working in a data-driven way, you can often get other data at your disposal, so that the result can be completely different from what it was expected (Director I/Principle manager, Paragraph 8).

Multidisciplinary teams, knowledge sharing, and teamwork were also mentioned as the positive aspects of an Agile way of working*:*

It's not needed that everyone has the same level of knowledge in the team. One member is an expert and has a certificate. The other members let him guide in that area (Interview transcript- BI-consultant, Paragraph 247).

Use of data development process

Within different projects, the experts mentioned different data development process approaches. There is a clear understanding that developing a data product is different than developing software. The principal manager said that "when we look at data-driven projects, the process of developing a data product is slightly different than with software development" (Transcript Director 1- Principal Manager, Paragraph 25).

All the experts state that delivering a minimum viable product that meets its stakeholders' requirements is a great challenge for the teams in data-driven projects. They mentioned that defining an MVP by a business and focusing on delivering the product is the challenge for their teams. The director said, "Always think from the customer's perspective, so don't start in technology" (Transcript Director 2-Principle Expert - Paragraph 11).

Use of Agile Methodologies

An Agile team aims to iteratively and incrementally deliver a minimum viable product. Within the projects that the experts are working on, Scrum is the Agile methodology that is used for developing data products. The journey of developing a product begins with an idea, question, or problem for the business. The director said, "Always look at what interest do business has, what does the customer want?" (Transcript Director 2-Principle Expert, Paragraph 9). He describes the journey as follows:

A question from a business, that's when everything gets started. The business picks up someone, an information analyst, or usually a business consultant to describe the question. You are far from thinking in a project yet. The next step that needs to be taken is to do a small analysis; what does the business question mean and what can be delivered already with what I have? What data do I have available and can I answer the question with this data? So, you first look at that data. Then you go and see what data is missing; do I have to retrieve external data or is that data not there or do I have

to combine things? In short, what can I do, what do I need? Then you finally come to an approach and then you start thinking about how you could report it. I think if you have an overview of that process, you will only see which team I need. And what technology do I have available?

The roadmap that is described by the director seemed closer to Waterfall than Agile. Although there is no question that there is iterative and incremental development in his roadmap, he emphasized his opinion about Scrum as follows: "The important thing is that you realize Scrum is not a black box" (Transcript Director 2-Principle Expert, Paragraph 57). Scrum or any other Agile framework is created to support the Agile teams to work with transparency and deliver high-quality products frequently.

Use of Agile best practices

Some of examples of Agile practices that experts mentioned during their interviews include trust between team members to achieve the common goal (multidisciplinary team), a clear understanding of the business problem (refinement), regularly showing progress (iterative and incremental development process), continuous improvement, automated testing, continuous deployment, and just in time designing and architecting.

Agile team

The director I-principle manager stated that

> *to create an Agile team, every organization needs to take some decisive and critical decision in managing the organization. In an Agile project environment, the business is involved in the development process. It's important that business and upper management act based on equality in the project organization and release using the power. An Agile organization wants to be an eco-system where everyone, with each other, is connected in any way (Paragraph 7-62).*

He also stated that "establishing trust and retaining the structure that team members establish with each other" are the first steps that need to be taken by an organization to support its Agile teams.

Director 1-principal manager stated that within an Agile team, "craftsmanship is one of the basic conditions" (Paragraph 76). Defining position help the Agile team trust each other in their daily work and the responsibilities that they have. Experts mention the following positions in an Agile team: the product owner responsible for defining and describing the minimum viable product; the Scrum master/Agile coach who supports the team in applying Agile during the development process; the product team, which is responsible for making and delivering the data product and may consist of data engineers, software engineers, BI consultants, testers, business analysts, and information analysts; and domain experts, platform engineers, and solution architects, who are supporting teams with their specialties during the development process.

Design verification

Once a draft design was ready, the elements of the framework (development method, development cycle, Agile practices, and stakeholders) are presented to the experts during a validation presentation.

Validation presentation

During this presentation, the following experts who are interviewed were present: the BI developer/Product owner, BI consultant/Scrum master, data/solution architect, and the tester. Table 5 presents an overview of the modifications that were made in design after this presentation.

Table 5: Results of design verification

		Before the verification presentation	After the verification presentation
AFDS Structure	Development Method	The way that the process was visualized was not fully understandable. The actions were not clear, and the direction of the development cycle was not correct. Aligned the direction of the arrows. Placed the inner circle on top of the outer arrow so it looks like one. Changed colors for better contrast. The first design has been included in Appendix VIII-AFDS first concept, Figure 24.	The development process was adjusted; the visualization was improved based on the received feedback. Figure 16: AFDS Development method illustrates the AFDS development.
	Development Cycle	There were discussions about the generality of the cycle for developing various types of data products. Choosing different names for activities and improving the visualization was the feedback that could improve the cycle according to experts. The first design has been included in Appendix VIII-AFDS first concept, Figure 24: First design of AFDS development method.	Based on the received feedback, the development cycles were refined. Some activities were renamed, and the cycle visualization was adjusted. Figure 15 illustrates the AFDS development cycle.
	Agile practices	The chosen Agile practices were enough.	No changes
	Stakeholders	There were some discussions about the grouping of the stakeholders and the roles of customers and the product team. The first design has been included in Appendix VIII-AFDS first concept, Figure 25.	The grouping of the stakeholders was modified, and the roles were refined. Figure 18: AFDS project illustrates AFDS project stakeholders.

6.2.Summative evaluation - Evaluation of demonstration

This section contains the evaluation of the results of the AFDS demonstration. The AFDS framework as the design science artifact is evaluated based on the evaluation criteria of Prat et al. (2014) described in section Evaluation:

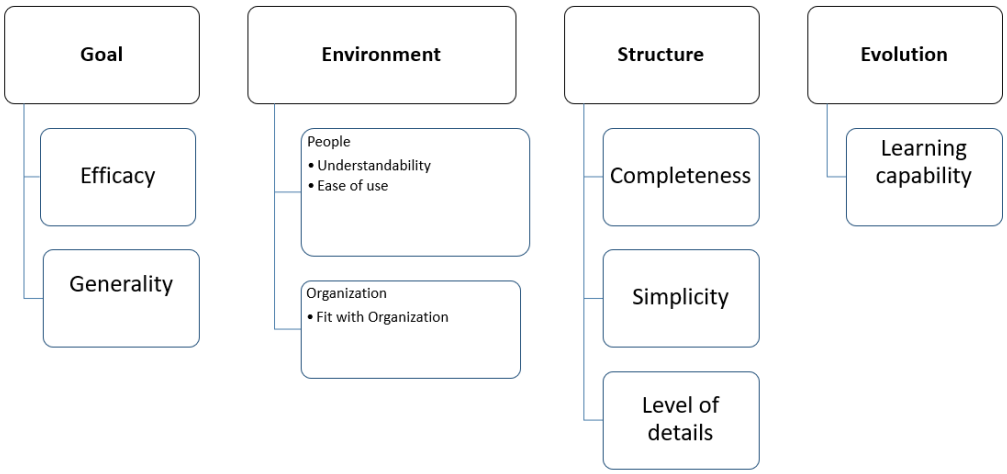


Figure 21: Criteria for AFDS validation

Efficacy is the degree to which AFDS provides its desired effect based on the requirements. Generality provides the framework to be applied in a broader project area such as data science, data analytics, business intelligence, big data, and linked data. People and organizations are chosen as evaluation criteria from the environmental dimension to verify the consistency of the framework. The structure of AFDS is evaluated based on its completeness, simplicity, and level of detail. The learning capability from the evolution dimension is used to evaluate the capacity of the framework from the experience and reactions of the environment (Prat et al., 2014).

Due to the time limitation, it was unfortunately impossible to evaluate the accuracy and performance criteria of the framework. However, during the demonstration presentation and hands-on workshop, all the elements of the framework and the framework as a whole was validated by 1) working on a fictional use case from practical experiences of the experts and 2) an online survey answered by the attendees after the demonstration session. The results of the evaluation are explained and illustrated using analyses of the presentation, workshop, and online survey.

In the subsections below, the results of the validation are described based on the evaluation criteria which are illustrated Figure 21: Criteria for AFDS validation.

Goal

The experts indicated that they are willing to introduce the framework to their clients. The elements of the framework fit their clients' organizations and can improve the Agile implementation at their projects. Consultants determined that AFDS can be applied in data-driven projects such as business intelligence, big data, linked data, and data science. This contributes to the generality of the framework.

However, one of the main concerns of the consultants was including all disciplines on one team. The nature of work for a data scientist seems to be significantly different than that of the other project roles. The work is more exploratory, and research based, making it difficult to combine with the other disciplines that work together on a concrete data product.

Environment

By applying the AFDS during the hands-on workshop, the understandability and the ease of use of the framework are validated. The results indicated that the framework is easy to understand by all disciplines in a project. The elements of the AFDS are easy to use, and the framework fits in the company.

Structure

The results of the workshop validate that the level of detail was enough to be able to apply the framework. The experts indicated that they are generally satisfied with the AFDS development method, development cycle, Agile best practices, and stakeholders. However, the experts had doubts about the completeness of the framework and whether the framework provides enough guidelines for implementing and applying Agile. Applying the AFDS framework in a real-world environment and practicing the Agile way of working could create more trust within the organization about the applicability of the framework.

Evolution

Different discussions during the demonstration, indicate that a lack of knowledge about the data development processes and methodologies, in combination with a lack of knowledge and experience of the Agile principles and mindset, cause doubts for some of the consultants in applying the AFDS framework in their projects. This may cause continuous discussions, which will interfere in applying the framework successfully. The results from analyzing the video records indicate that, although there were doubts during the presentation, at the end of the workshop, the participants fully understood the elements and they were pleasantly surprised about the usefulness and ease of use of the AFDS framework. The director was convinced that they can introduce AFDS to one of their clients to improve their project development pipeline.

7. Conclusions and recommendations

This chapter describes the conclusions that are drawn from the results of the design science research presented in this thesis. The conclusion is followed by the recommendations for the organization and for future research. The chapter concludes with a personal reflection on the results and the process of performing the DSR for this thesis.

7.1. Conclusions

Based on the adaptive nature of Agile, it should be suitable for nearly any project; however, this requires a deeper understanding of its principles and values. The challenge that remains for data-driven projects is the adoption the Agile practices and methodologies that most effectively facilitate data-driven projects. To address this problem, the DSR of this thesis aimed to answer the main research question:

How can Agile principles, practices, and methodologies be applied in data-driven projects to improve and effectively facilitate these projects?

The theoretical framework (displayed in Figure 6) provides guidelines with a focus on Agile data development method and Agile practices for data-driven projects and answers the sub-questions of the research.

The DSR project aimed to develop a framework that supports data-driven organizations in applying Agile in their projects. To achieve this goal, the design science research framework of (Hevner et al., 2004) is used as a guideline to create a framework that is the design artifact for this research.

7.2. Research implications

The DSR of this thesis identifies and structures three contributions to the field of information science.

First, the literature review considered scientific literature; analyzed Agile principles, practices, and methodologies within the data-driven field; and identified a theoretical framework that can be useful for future design science research.

Second, based on the theoretical framework, the AFDS framework as the designed artifact of DSR provides a set of guidelines for adopting Agile in data-driven projects.

The data development cycle of AFDS, which is derived from the CRISP-DM process model and fast analytics/data science lifecycle, enables multidisciplinary teams to develop data products quicker and more effectively. Due to the time limitation, the use of this development cycle by data-driven teams was not investigated. By conducting a qualitative study, the AFDS data development cycle can be investigated, and it becomes possible to determine how effective it is for data-driven teams.

Third, the AFDS framework uses the Scrum guidelines and process combined with the development cycle, which are appropriate for data-driven projects. The results indicate that the Scrum process is suitable for data-driven teams. A qualitative study can investigate if applying the AFDS framework in a data-driven team that already works with Scrum can improve team performance.

Fourth, the AFDS framework is designed and evaluated in an organization that desires to transition from a traditional data development process to a dynamic Agile approach. A qualitative study can further investigate and validate the applicability of the framework in a data-driven organization that desire to move from classic data development approaches to Agile.

7.3. Practical implications

The observation and validation of the designed framework are conducted using a questionnaire, five semi-structured interviews, a validation presentation, a demonstration by a presentation/workshop, and an online survey. The results of DSR illustrate that lack of knowledge about data development cycles in combination with a lack of experience with the Agile way of working, which causes major discussions that interfere with the implementation of the framework in projects. This makes it difficult for the company to introduce the framework to its clients. Therefore, it is recommended that the company starts to organize hands-on workshops and training for its consultants to improve their knowledge and experiences about data development approaches, adopting Agile in their daily work. This would allow experts to learn and create a better understanding of the effects of the framework in their projects. When the consultants are more aware of the content of AFDS and know how to apply it in their project, it could help them to share their experience more convincingly with their clients and suggest them to apply AFDS to their projects.

In order to apply the framework to a data-driven project, it is recommended that the organization provide training, workshops, and meet-ups to the members of the project and teams to help them become acquainted with the framework and understand the advantages of it to their projects.

7.4. Limitations and future research

The most important limitations of this study include the following:

- There were only five consultants from the company that were involved in the field observation, evaluation, and implementation of the framework. This low number of experts is identified as a limitation and a threat to the validity and quality of the research. The possibility exists that the consultants, if biased, impact the quality of the evaluation. Therefore, the involvement of a higher number of experts in the formative and summative evaluations could reduce the possible impact of being biased.

- Although, all experts had different roles in their projects, unfortunately, it has not yet been possible to interview other possible stakeholders and people in different positions. It would be useful to conduct similar research with more interviewees from different positions in the project. If more participants were available during design and development, more feedback could be used in designing the framework.

- The AFDS framework as the DSR designed artifact of this thesis has not been implemented in the real world and therefore has not been fully evaluated. Nevertheless, the implementation was performed through a hands-on workshop with the consultants of the company and the evaluation is performed in a rigorous way, which is described in chapter 6. To support a real implementation of AFDS, additional proof is needed that the framework is valid and correct.

The limitation describes the above result in the following recommendations for future research:

- Future research is needed to implement the designed artifact of this DSR (AFDS framework) in one or more organizations that desire to apply Agile in their data-driven projects to evaluate the implementation of the framework in the real world.

- The implementation of the framework in various kinds of data-driven projects (e.g., data science, business intelligence, and big data) is recommended. The research needed to focus more on the applicability of the framework to different data-driven project directions.

7.5. Reflection

At the start of the research, there was not much awareness about the design science research and how differs from empirical research. I had not earlier worked with this methodology, and it was my first time applying it in my research. Despite the proposed literature by my supervisors about DSR, it was quite challenging to understand the aim of the methodology and apply it correctly for the first time. By performing the project research, the aim of the methodology and the advantages of DSR in the field of information science became increasingly clearer. By applying DSR, which eventually means building and evaluating IT artifacts for solving business problems in the field of IS, the academic research is compatible and useful for the field of information science and technology. The AFDS framework as the designed artifact of this DSR may be used and directly applied as the solution to the business and practical problems. I believe this is the most significant advantage of DSR in comparison to classical research methodologies.

The topic of the research, "Data Science goes Agile," was very interesting for me. I was aware that my expertise as an Agile IT consultant in software engineering projects could enable me to analyze the problem both in breadth (Agile) and in-depth (data science). When the literature review was completed, I had a good understanding of the problem and could easily define the objectives of my research. During the literature review, it was striking that data science, data analytics, data mining, big data were used interchangeably in various literature. During the research project, I understood that there was also confusion about the definition of a data science project and the role of a data scientist. The results of the evaluation indicate that the framework could be used in a broader context than just in data science. I have therefore decided to change the name of the framework to "Agile Facilitate Data-Driven Solutions" to generalize the applicability of the framework in a broader context.

Since I was not familiar with the company where I performed my research, it was quite exciting to manage the planning and monitor the scope. The consultants of the company were working fulltime on different projects located all over the Netherlands. It took me almost three weeks to contact a group and plan the expert interviews. From the beginning, I was aware of the risk that the research could not be completed within the defined time frame. Continuously monitoring the planning and the scope was needed to be able to complete the research in the defined time frame. I believe that tight planning helped me to manage the process, to stick to the agreement, and be able to complete the research without delay.

Due to the schedules of the consultants, I had to plan the interviews over a few days close to each other. Three experts were interviewed in one afternoon, and the others were interviewed over two consecutive days. This prevented me from having enough time to reflect and evaluate the interviews and control the data to prevent issues and improve the quality of the material. When I was transcribing the audio material and coding the text to analyze the data, I realized that there were some tools and techniques that I could use to be able to process the data faster. For example, using "Google Speech Recognition" could help me to transform the audio materials to text for the transcription process, which I was the most labor-intensive part of the research. After the transcription of two interviews, I started to use audio transcription tools such as "f4transkript" and "f4analyse," which helped me to process data more efficient and be able to analyze them faster. I would use these tools for coding procedure again in the future.

As it was not the first time for me to conduct semi-structured interviews, I was convinced that I had gathered all the necessary data from the interviews. But in retrospect, during the transcription, I realize that if I had done better, I could have retrieved higher quality data. There were times where I

wondered why I had talked so much instead of letting the experts express their ideas. In one interview, there was a strong tendency to fill in answers for the expert as I am experienced in Agile software development. I realize that I should be more patient and give as much space as possible to the interviewees/experts and allow them to talk. In future research, I will plan such interviews over a certain amount of the time. I will use technologies and techniques for recording the interview, which helps me to transcribe them more efficient. And finally, I will transcribe the interview directly after recording or within the 24 hours. Then, I will be able to evaluate the data faster and improve the followed interviewed.

References

Ambler, S. . (2012). Agile Modeling (AM) Home Page: Effective Practices for Modeling and Documentation. Retrieved November 25, 2018, from http://agilemodeling.com/

Anderson, C. (2015). Chapter 1. What Do We Mean by Data-Driven? In *Creating a data-driven organization* (p. 285). Retrieved from https://www.oreilly.com/library/view/creating-a-data-driven/9781491916902/

Asay, M. (2017). 85% of big data projects fail, but your developers can help yours succeed. Retrieved November 8, 2018, from TechRepublic website: https://www.techrepublic.com/article/85-of-big-data-projects-fail-but-your-developers-can-help-yours-succeed/

Batarseh, F. A., & Gonzalez, A. J. (2018). Predicting failures in agile software development through data analytics. *Software Quality Journal*, *26*(1), 49–66. https://doi.org/10.1007/s11219-015-9285-3

Batra, D. (2018). Agile values or plan-driven aspects: Which factor contributes more toward the success of data warehousing, business intelligence, and analytics project development? *Journal of Systems and Software*, *146*, 249–262. https://doi.org/10.1016/j.jss.2018.09.081

Beck, K., Beedle, M., Bennekum, A. van, Cockburn, A., Cunningham, W., Fowler, M., … Thomas, D. (2001). Principles behind the Agile Manifesto. https://doi.org/10.1002/mrdd.20074

Beck, K., Beedle, M., Bennekum, A. van, Cockburn, A., Cunningham, W., Fowler, M., … Thomas, D. (2014). Manifesto for Agile Software Development. Retrieved September 17, 2018, from 2001 website: http://agilemanifesto.org/

Cao, L. (2017). Data Science: A Comprehensive Overview. *ACM Computing Surveys*, *50*(3), 1–42. https://doi.org/10.1145/3076253

Chan, F. K. Y., & Thong, J. Y. L. (2009). Acceptance of agile methodologies: A critical review and conceptual framework. *Decision Support Systems*, *46*(4), 803–814. https://doi.org/10.1016/j.dss.2008.11.009

Cohn, M. (2004). *User stories applied : for agile software development*. Addison-Wesley.

Dawson, C. S., Hannigan, K. E., Hattrup, G., & ... (2011). Apparatus, system, and method for transparent end-to-end security of storage data in a client-server environment. *US Patent …*. Retrieved from https://patents.google.com/patent/US7899189B2/en

Dhar, V. (2012). *Data Science and Prediction*. https://doi.org/10.2139/ssrn.2086734

Fowler, M. (2004). Writing The Agile Manifesto. Retrieved September 17, 2018, from Martinfowler.com website: https://martinfowler.com/articles/agileStory.html#TheSnowbirdMeetingAndTheManifesto

Hevner, A. R. (2007). *A Three Cycle View of Design Science Research A Three Cycle View of Design Science Research*. *19*(2), 87–92.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105.

Hoogendoorn, S. (2014). *This is Agile: Beyond the basics. Beyond the Hype. Beyond Scrum.* Dymaxicon.

Jurney, R. (2017). *Agile data science 2.0 : building full-stack data analytics applications with Spark*. Retrieved from http://shop.oreilly.com/product/0636920051619.do

KDnuggets.com. (2014). CRISP-DM , still the top methodology for analytics , data mining , or data science projects. Retrieved November 12, 2018, from KDnuggets website: http://www.kdnuggets.com/2014/10/crisp dm top methodology analytics data mining data science projects.html

Krawatzeck, R., Dinter, B., & Thi, D. A. P. (2015). How to make business intelligence agile: The agile BI actions catalog. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2015-March*, 4762–4771. https://doi.org/10.1109/HICSS.2015.566

Kretschmer, R., Pfouga, A., Rulhoff, S., & Stjepandić, J. (2017). Knowledge-based design for assembly in agile manufacturing by using Data Mining methods. *Advanced Engineering Informatics*, *33*,

285–299. https://doi.org/10.1016/j.aei.2016.12.006
Kuechler, B., Petter, S., & Esearch, S. Y. R. (2004). *Design Science Research in Information Systems*. (March), 1–66. https://doi.org/10.1007/978-3-642-29863-9
Larson, D., & Chang, V. (2016). A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science. *International Journal of Information Management*, *36*, 700–710.
Li, Y., Thomas, M. A., Osei-Bryson, & Muata, K. (2016). A snail shell process model for knowledge discovery via data analytics. *Decision Support Systems*, *91*, 1–12. https://doi.org/10.1016/j.dss.2016.07.003
Marbán, O., Segovia, J., Menasalvas, E., & Fernández-Baizán, C. (2008). Toward data mining engineering: A software engineering approach. *Information Systems*, *34*(1), 87–107. https://doi.org/10.1016/j.is.2008.04.003
Mariscal, G., Marbán, Ó., & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *Knowledge Engineering Review*, *25*(2), 137–166. https://doi.org/10.1017/S0269888910000032
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *International Journal of PharmTech Research*, *1*(4), 1685–1691. https://doi.org/10.1371/journal.pmed.1000097
Nascimento, D., Santana, G., & de Oliveira, A. A. (2012). *An Agile Knowledge Discovery in Databases Software Process*. (c), 56–64. https://doi.org/10.1007/978-3-642-34679-8_6
Okoli, C., & Schabram, K. (2011). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *Sciences-New York*, *10*(2010). https://doi.org/10.2139/ssrn.1954824
Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, A. S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 45–77. https://doi.org/10.2753/MIS0742-1222240302
Prat, N., Comyn-Wattiau, I., & Akoka, J. (2014). Artifact evaluation in information systems design-science research. *18th Pacific Asia Conference on Information Systems*, 16. Retrieved from http://aisel.aisnet.org/pacis2014
Saltz, J., Shamshurin, I., & Crowston, K. (2017). Comparing Data Science Project Management Methodologies via a Controlled Experiment. *Proceedings of the 50th Hawaii International Conference on System Sciences*, 1013–1022. Retrieved from http://hdl.handle.net/10125/41273
Saltz, J., Wild, D., Hotz, N., & Stirling, K. (2018). *Exploring Project Management Methodologies Used Within Data Science Teams*. 1–5.
Saunders, M. N. K., Lewis, P., & Thornhill, A. (2012). *Research methods for business students*. Pearson.
Schmidt, C., & Sun, W. N. (2018). Synthesizing Agile and Knowledge Discovery: Case Study Results. *Journal of Computer Information Systems*, *58*(2), 142–150. https://doi.org/10.1080/08874417.2016.1218308
Schwaber, K., & Sutherland, J. (2011). *The Scrum guide*. *2*(July), 17. https://doi.org/10.1053/j.jrn.2009.08.012
Sekgweleo, T. (2015). *Understanding Agile System Development Methodologies*. *5*(7), 18–24.
TechTarget. (2017). What is Best Practice? Retrieved November 18, 2018, from https://searchsoftwarequality.techtarget.com/definition/best-practice
The Python Standard Library. (2016). Unit testing framework — Python documentation. Retrieved November 25, 2018, from https://docs.python.org/3/library/unittest.html
Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Appendix I- Literature review procedure

A lack of knowledge and experience about Agile in organizations causes difficulties in implementing Agile in their data-driven projects (Li, Thomas, Osei-Bryson, & Muata, 2016). To gain a comprehensive understanding of this problem, the three sub-questions below were formulated in Chapter 1 for this research.

SQ1: What is the development cycle of an Agile data-driven project?

SQ2: Which Agile principles, practices, and methodologies that are known and being used in data-driven projects?

The literature review aims to answer these sub-questions and generate knowledge (a theoretical overview) to support the design and development of the artifact for this research.

Planning

Creating a starting set

Before beginning to develop the search, a starting set of relevant references was created that helped identify what concepts and keywords can be incorporated into the search strategy. There were nine articles recommended by the research supervisors regarding the research theme "Data Science Goes Agile." These nine articles are displayed in Table 6; they formed the starting set of this research. This collection was used to identify the concepts need to search for the articles for the literature review.

Table 6: Starting set for literature review

Article #	Description
A1	Do Nascimento, G. S., & De Oliveira, A. A. (2012). An Agile Knowledge Discovery in Databases Software Process
A2	Li, Y., Thomas, M. A., & Osei-Bryson, K. M. (2016). A snail shell process model for knowledge discovery via data analytics.
A3	Mariscal, G., Marbán, Ó., & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies.
A4	Saltz, J., Heckman, R., Shamshurin, I., & Saltz, J.; (2017). Exploring How Different Project Management Methodologies Impact Data Science Students.
A5	Saltz, J., Wild, D., Hotz, N., & Stirling, K. (2018). Exploring Project Management Methodologies Used Within Data Science Teams.
A6	Saltz, J. S., Shamshurin, I., & Crowston, K. (2017). Comparing Data Science Project Management Methodologies via a Controlled Experiment.
A7	Schmidt, C., & Sun, W. N. (2018). Synthesizing Agile and Knowledge Discovery: Case Study Results.
A8	Sharma, V., Stranieri, A., Ugon, J., Vamplew, P., & Martin, L. (2017). An Agile Group Aware Process beyond CRISP-DM.
A9	Shen, B. (2018). Universal knowledge discovery from big data using combined dual-cycle.

Identifying key concepts

The first three sub-questions of this research and the information discovered from the starting set resulted in the list below, which was created to determine key concepts and terms for researching the literature for study:

1- Data-driven
2- Lifecycle
3- Agile
4- Methodologies
5- Project

Each of these words was used as a concept, and for each concept, a list of terms was derived from the starting set or the references mentioned in the articles from the staring set (snowballing).

Combining key concepts/terms

The two Boolean operators “AND” and “OR” were used to combine the terms (OR) from different search concepts (AND).

Table 7: Combining key concepts/ terms

	Key terms (AND)				
	Data - Driven	Development cycle	Agile	Methodologies	Project
Related terms OR	• Big Data • Data science • Data mining • Data analytics • Knowledge discovery	• Process	• Software development • Principles • Practices • Implications	• Models • Methods • Frameworks	• Team • Stakeholders • Implications

The Boolean operations were not used in Google Scholar because this search engine was employed for a wide-ranging search. Quote were used to search for exact terms in Google Scholar.

Table 8 displays the queries that were executed in the search engines.

Table 8: Executed search-queries

	Google Scholar	Results
Queries	"data science" "Agile principles" "projects" since 2014	46
	"data mining" "Agile best practice" Since 2014	7
	"Agile" "data science projects" "stakeholders" since 2014	37
	"Implication" "Agile" "data mining projects" since 2014	12
	OU Library	**Results**
Queries	((TitleCombined:(Data)) AND (Agile) AND (Fulltext:(Transformation);	174
	(TitleCombined:(Data science)) AND (TitleCombined:(Agile)) AND (Fulltext:(Projects))	9
	(((TitleCombined:(Agile)) AND (Abstract:(Agile principles)) OR (Abstract:(Agile methodes)) OR (Abstract:(Agile guidlines)) OR (Abstract:(Data science))) AND ((TitleCombined:(Data science)) OR (TitleCombined:(Data mining))) ;)	13
	(TitleCombined:(Agile)) AND ((TitleCombined:(Data science)) OR (TitleCombined:(Data mining))) AND ((Agile methods) OR (Agile principles) OR (Agile guidlines))	9
	(TI agile) AND (TI data science) OR (TI data mining) AND (project team) Criteria: Published Date: 20130101-20191231Scholarly (Peer Reviewed) Journals Language: English	25
	TI Data science AND TI agile	3

The results of each query were sorted from newest to oldest.

Search engines

The digital library of the Open Universiteit and Google Scholar were used as search engines for the following reasons:

- Google Scholar was used for wide-ranging searches to discover a variety of material and quickly explore related articles that could be noteworthy. It was useful because it revealed the number of times an article had been cited and by whom.
- The advanced search at the Open Universiteit library made it possible to execute a Boolean search in the world's largest digital libraries in the field of information technology.

Selection

Searching began by running the queries in the aforementioned search engines. The goal of this step was to create a list of at most 500 articles and properly select articles from this list based on the general criteria. There were three general search criteria used to limit the search result:

- Peer-reviewed articles: The sources must be peer-reviewed articles to use them for academic research.
- Date >= 2012: The timeline was determined to limit the search result and ensure the articles are recent. The result could not be published prior to 2012.
- Keywords: The title of an article should contain at least two key terms: "Agile" and "Data science."
- Language: English
- Discipline: Computer science and software engineering were used to better specify the search result.

Extraction and execution

A PRISMA flowchart, illustrated in Figure 22, was used to analyze, review, and refine the collection from search results and create a final set for the literature review (Moher et al., 2009).

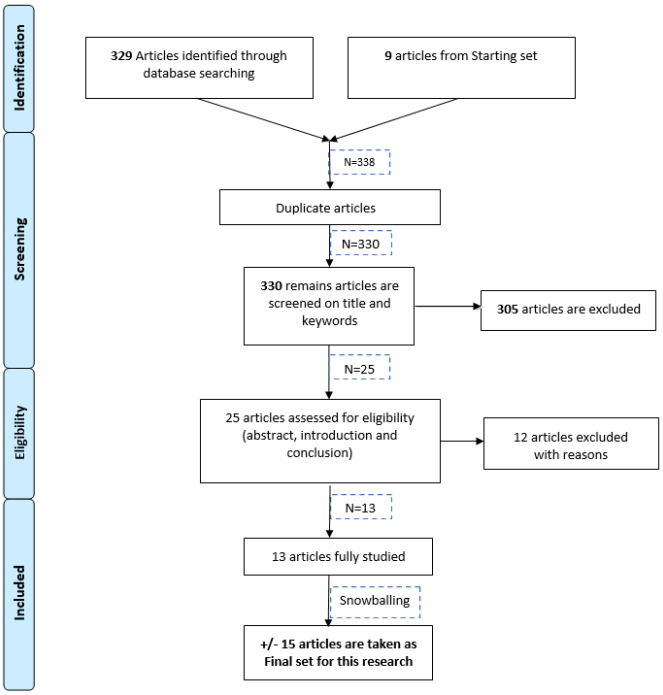


Figure 22: PRISMA flowchart (Moher et al., 2009)

Identification:

From the database searches, 329 articles were identified. Including the nine articles from the starting set, 338 articles were used to begin creating a relevant set for the literature review.

Screening:

First, duplicate articles from the two sets were removed. Most of the articles from the starting set, recommended by the research supervisors, were discovered in the database searches. Duplicate articles were removed, and the 330 articles that remained were candidates for the subsequent screening step. The following step was to examine the 330 articles and exclude those that did not appear to be suitable for this research. The subject of the title of these articles was screened. The subject needed to be related to the theme "Data science goes Agile." Some articles were related to the key concepts, but their subjects revealed that they were not appropriate for this research. In total, 305 articles were excluded after this screening, resulting in 25 articles that proceeded to the next step.

Eligibility:

This step aimed to discover the articles and information that were more relevant to the research questions. The information in the "Abstract," "Introduction," and "Conclusion" of these articles had to be related to the research questions. As a result, 12 articles were excluded, and 13 articles were finally selected for the literature review of this research.

Included:

Finally, the 13 remaining articles were thoroughly read and used as references for the literature review of this research. The snowballing method was used with the 13 articles in the final set that have references relevant to this research but were not included in the search results. The final set of the articles consisted of 15 articles used for the literature review.

Appendix II- Interview protocol

Interviewee name:
Interviewee role:
Interviewee experience:
Interview timestamp:
Interview duration:

Objectives to retrieve from interviews:

- How is Agile organized in the project?
- Which Agile methodologies, techniques and practices are used in the project?
- What are the requirements and constraints for an Agile framework?

Introduction:
Have a small introduction talk with the interviewee to get to know each other. Make a brief summary of the research questions and the steps that are taken to get answers to the questions.

Context / background
The paragraph below explains shortly what the research is about:

> "The subject of my research is "Agile Facilitates Data Science". The aim of this research is to design an Agile framework for data driven projects / teams. The framework addresses the problems that emerged from previous studies within the data driven project and offers an iterative and incremental development process together with a number of guidelines (Agile practices) for Data Science projects and teams to improve their performance".

The paragraph below explains how the interviewee fits in this research:

> "There are different roles defined in an Agile team. From your role as "…" I would like to know how Agile is organized at your client's project organization and what do you think influences the success of implementing Agile within data driven teams?"

Open question	**Probing question**
How is Agile organized in your team?	Which framework, methods and practices are used?
How does Agile influence your team?	
Which data process approaches does your team use to develop a data product?	CRISP-DM, KDD or some other approaches?
Which roles does your team consist of?	Is there a data scientist in the team?
Which Agile practices are used by your project?	TDD, CI, CD, Pair programming, Agile Modeling?
Which tools and techniques does your project use to support Agile implementation?	Travis CI, Circle CI, Jenkins?
What do you think about the influence of Agile within Data driven projects?	
What do you think influences the success of implementing Agile within data driven teams?	

Appendix III- Questionnaire

How Agile is your team?

Thank you for agreeing to take part in this survey. It is being conducted as part of a research project aiming to design an Agile framework which facilitates data driven (data driven) teams.
All your answers and comments will remain confidential and anonymous. If you have any questions about the survey please contact Maryam Donker (maryam.donker@aylex.org).

Nr. Of questions: 30
*Required

1- What is your role in the project? (multiple answers possible) *

- ☐ Development team member
- ☐ Scrum master
- ☐ Team lead
- ☐ Product owner
- ☐ Delivery manager / Director
- ☐ Product / Program / Project manager
- ☐ Other:

2- Which disciplines are in your team? (multiple answers possible) *

- ☐ Data Scientist/ Data Analyst
- ☐ Tester
- ☐ Software engineers
- ☐ BI Developers
- ☐ Business analyst
- ☐ Domain experts
- ☐ Architect
- ☐ Platform engineers
- ☐ Scrum master
- ☐ Product owner
- ☐ Other:

3- Which Agile framework does your team use? (select one) *

- ○ Scrum
- ○ Kanban
- ○ Company-Owned
- ○ None
- ○ Other:

4- Which data process approaches do your team use to develop a data product? (select one) *

- ○ KDD (Knowledge Discovery in Data)
- ○ CRISP-DM (Cross Industry Standard Process for Data Mining)
- ○ Other:

5- How often is your team able to deliver a (minimum) viable product? *

- Never
- Occasionally
- Sometimes
- Often
- Always

6- How often do you think your team delivers what the business needs most? *

- Never
- Occasionally
- Sometimes
- Often
- Always

7- Is your team accommodating all the changes suggested by the product owner? *

- Yes
- No
- Other:

8- How often does your team have insight into what it will be doing in the next iteration? *

- Never
- Occasionally
- Sometimes
- Often
- Always

9- How often are the work items that enter the current iteration, ready to be implemented? *

- Never
- Occasionally
- Sometimes
- Often
- Always

10- How often is your team being disrupted and controlled by outsiders during the iteration? *

- Never
- Occasionally
- Sometimes
- Often
- Always

11- Is your team collaborating daily? (Ex: Daily stand-up) *

- Yes
- No
- Other:

12- Does your team visualize the workflow within each iteration? E.g. Scrum/Kanban board *

- Yes
- No

- Other:

13- What are the steps of your team's workflow within an iteration? *

14- Is your team motivated to deliver the agreed deliverables at the end of each iteration? *
- Yes
- No
- Other:

15- Do your project team members trust each other? *
- Yes
- No
- Other:

16- How often, when it's possible, does your team communicate face to face? *
- Multiple times a day
- Once a day
- Couple times per week
- Once a week
- Couple times per iteration
- Once per iteration

17- How often is your team able to deliver a fully tested, working product at the end of the iteration? *
- Never
- Occasionally
- Sometimes
- Often
- Always

18- Does your team have a sustainable pace? *
- Yes
- No
- Other:

19- Is your team rushing towards deadlines? *
- Yes
- No
- Other

20- Are the team members locked into their specific roles? (They can/will do only one type of task) *
- Never
- Occasionally
- Sometimes
- Often
- Always

21- How often is your team able to reflect on whatever has happened? (Ex: Retrospective) *
- Daily

- Every iteration
- Every few iteration
- Hardly
- Never

22- Does your team take appropriate actions based on whatever has come up in retrospectives? *
- Never
- Occasionally
- Sometimes
- Often
- Always

23- Is your team continuously improving what it does? *
- Yes
- No
- Other:

24- Which of these best practices are used by your team members? (multiple answers possible) *
- ☐ Continues active participation and collaboration of stakeholders
- ☐ Pair programming
- ☐ Agile Modeling
- ☐ Just in Time (JIT) documentation,
- ☐ Test Driven Development (TDD)
- ☐ Continuous testing
- ☐ Automated testing
- ☐ Continuous deployment
- ☐ Continuous integration
- ☐ Continuous architecting
- ☐ Continues improvement
- ☐ Other:

25- Which of above Agile practice(s) do you consider to be important for a data driven team. *

26- What do you think about the influence of Agile within Data driven projects? *

27- How do you think Agile can influence data driven teams? *

28- How does Agile influence your team? *

29- How is Agile organized within your team? *

30- What do you think influences success of implementing Agile within data driven teams? *

Appendix V- Results of expert interviews

Table 9: Codebook

Code	Text
Agile way of working	"Wij hadden tot op kort, sprints van 4 weken. Die zijn nu verkort naar 3 weken. Daar zijn we niet helemaal blij mee. De overhead neemt dan toe. Je moet toch gewoon de sprint planning doen, de retrospectie en dat soort zaken. Maar dat neemt relatief veel tijd in beslag binnen een sprint. Of je nu 3 of 4 weken doet. Die planningssessie is er, en die retrospectie is er. Dat kost relatief veel tijd. Als je kijkt naar doorlooptijd van de rest van de sprint. We doen dus de planning, dan wordt gekeken naar wat is er in de vorige sprint niet afgekomen. Dat komt weer bovenaan de backlog van de volgende sprint. Dan wordt er gekeken naar welke punten, welke stukken functionaliteit kunnen we verder oppakken. De PO geeft dan volgorde aan. Dan zou je ook nog moeten gaan pokeren. Kijken naar, past het er allemaal in. We doen niet aan klassiek pokeren met de Fibonacci reeks. Dat hebben we in het begin 1 keer gedaan, maar daar zijn we weer vanaf gestapt. Er wordt nu meer gekeken naar de hoeveelheid tijd dat iets kost." (Transcript-Tester, Paragraph 49)
	"Agility betekent wendbaarheid," (Trsanscript- Directeur 2-Principle Expert, Paragraph 8)
	"ik ben wel voorstander om niet met hele dikke plannen te gaan werken, omdat je natuurlijk eigenlijk het einddoel, dat verandert." (Transcript- Directeur 2-Principle Expert, Paragraph 8)
	"Zeker als je data driven gaat werken, omdat je natuurlijk heel vaak andere data tot je beschikking kan krijgen, waardoor het eindresultaat totaal anders kan worden." (Transcript- Directeur 2-Principle Expert, Paragraph 8)
	"Altijd kijken met welk belang moeten wij, wat wil nou de klant?" (Transcript- Directeur 2-Principle Expert, Paragraph 9)
	"En dat is wel de manier hoe met name data science projecten werken. En dat heeft niet alleen met de data te maken, maar ook met de techniek, die natuurlijk heel snel nieuwe mogelijkheden meebrengt." (Transcript- Directeur 2-Principle Expert, Paragraph 9)
	"Altijd vanuit de klant denken, dus niet in de techniek beginnen." (Transcript- Directeur 2-Principle Expert, Paragraph 11)
	"De techniek is het hulpmiddel om naar die vraag toe te komen, alleen kun je al zien onderweg kan die vraag veranderen en dan ben je natuurlijk flexibel om het aan te passen." (Transcript- Directeur 2-Principle Expert, Paragraph 11)
	"Dat iedereen meer een gelijk kennisniveau heeft. De 1 is heel goed, heeft een certificaat. De ander draait maar mee in het team en laat zich leiden. En die denkt ook, wat moet ik dan precies bij een refinement doen? Dat kun je ook niet voorbereiden als je zo de meeting in gaat. En ook dat de organisatie weet van het proces. Die staat hier vaak nog wat los van" (Interview transcript- BI-consultant, Paragraph 247)

Code	Text
Data development Process	“we kijken naar data driven projecten, dan loopt het proces van het ontwikkelen van een product toch net anders dan bij softwareontwikkeling.” (Transcript- Directeur 1- Principal Manager, Paragraph 25) “Dat zijn alle bronnen. Op dit moment gaat dat allemaal naar een datawarehouse en daar zitten allemaal dashboards op middels click view enzovoort. Waar we naar toe willen in de praktijk, zijn we aan het kijken dat we naar de cloud moeten. Tegelijk met dat we naar de cloud willen, zeg ik, zet deze data in een data lake omgeving. Op die data lake omgeving kun je bijv. analyses doen, want de data is daar nog vers, nog raw, onbewerkt. Dus onbewerkte data en tegelijk wil je heel vroeg in je proces op die data governance toepassen. Zoals wat betekend die data? Welke business rules zijn er? Je wil die data classificeren. Dus deze data wordt gelijker tijd geclassificeerd. Mijn idee is om daar een nieuw platform voor neer te zetten die de governance in de gaten houdt. Dat is een tool die ook in de cloud zit. Die heet Callibra, het kan een andere tool zijn, dit is een data governance tool. Hierin kun je allerlei relaties, modellen, enzovoort in importeren. En je kunt ook zeggen wie de eigenaar is.” (Transcript-Architect, Paragraph 24) “Dit is een data lake, dat is ruwe data.” (Transcript-Architect, Paragraph 26) “Hier heb je analyse mogelijkheden.” (Transcript-Architect, Paragraph 26) “Dus hierna ga je de data in een data warehouse stoppen.” (Transcript-Architect, Paragraph 26) “Op die manier kun je dus het proces vanuit het verzamelen, analyseren of modellen maken kwalitatief beter doorlopen.” (Transcript-Architect, Paragraph 36) “En waar zit de Data Scientist? Op welke gedeel van het process gaat het data science gedeelte zitten? R: Data Science zit op verschillende lagen. Die kan op de raw data zitten, maar die kan ook verbanden gaan leggen met bijv. flight informatie. Dus data die volgens Schiphol. Dus Data Scientist zit over de hele breedte. Hij kan ook kijken naar de historie.” (Transcript-Architect, Paragraph 69 - 70) “I: Verschillende business owners hebben verschillende dashboards. De Data Scientist zou dan in verschillende processen de data kunnen gebruiken. R: Ja klopt, eigenlijk wil je een boundary om je Data Warehouse, en Data Scientist gebruiken ook hun data virtualisatie op het Data Warehouse.” (Transcript-Architect, Paragraph 71 - 72) “zaken als inschatten is vaak heel lastig. Zeker voor een developer. Die weet niet wat de achterliggende complexiteit is. Je ziet in veel Scrum projecten wordt gepokerd. Er wordt gevraagd, complexiteit 1 – 5 en het wordt op 3 punten gepokerd. Maar de werkelijkheid is vaak toch anders en loop je tegen allerlei issues aan in je data etc.” (Transcript-Architect, Paragraph 83) “heel belangrijk is, om vanuit de architectuur en design, probeert tooling en technieken te gebruiken dat het proces ondersteunt om sneller en kwalitatief betere producten te kunnen leveren.” (Transcript-Architect, Paragraph 87)

	“dat we niet precies weten waar het eindpunt is. Dus dat is dan ook veel moeilijker planbaar.” (Transcript-Tester, Paragraph 14) “Je wilt altijd iets opleveren wat met business waarde.” (Transcript-Tester, Paragraph 63) “Nou is dat vaak in BI land heel lastig te doen, omdat je met een DWH te maken hebt, met verschillende lagen. Dat zie je ook in Data Science. Daar heb je allemaal stappen. Die stappen zitten anders in elkaar, maar ik zie daar parallellen. Je hebt de eerste laag dat gegevens overgehaald worden uit bronsystemen. Of ingelezen worden vanuit bestanden. Dan komt er een laag waarin historie wordt vast gelegd. Dan komt er een laag waarin gegevens in een vorm worden gepresenteerd die voor gebruikers prettig is om op te querien. Je moet die verschillende lagen door. Wat je vaak ziet, is dat met in 1 sprint het ontsluiten van de gegevens doet en het aanmaken van de historische laag, en dat de volgende laag, de data mart laag, de gebruiksvriendelijke laag. Dat dat pas weer een volgende sprint wordt opgepakt. Dus je levert eigenlijk niet in 1 sprint heel veel business waarde op. Dan zie je dus ook dat de domein deskundige, zit te klagen dat het zo lang duurt voordat hij iets krijgt.” (Transcript-Tester, Paragraph 63) “Je moet 1 bron ontsluiten. Wil je business waarde opleveren? Dan zou je wellicht 1 of 2 veldjes bijv. moeten oppakken voor de 1e laag, die moet doorzetten naar de volgende laag, de data vault laag bijv. En die weer moeten doorzetten naar de data mart laag. Dus dan ga je heel smal opereren. Dat is iets wat een beetje haaks staat op de oude praktijk binnen BI. Daarin ontsluit je de bron en dan neem je zoveel mogelijk gegevens mee die je eventueel later nog kan gebruiken. Je hoort altijd die ontwikkelaars zeggen, die velden hebben we nu eigenlijk nog niet nodig, maar we nemen ze mee want dat is wel zo makkelijk. De eerste laag is dat nog niet zo erg. Dat is redelijk 1 op 1, maar bij de volgende laag begint dat veel complexer te worden. En bij de daarop volgende laag is het nog veel complexer, want bij een DWH neemt de complexiteit van de gegevens, van de business rules toe, naarmate je verder in de pyramide van de gegevens gaat. Want de 1e laag heeft nauwelijks business rules, vrijwel nooit. De 2e laag, de raw vault van de data vault, is redelijk simpel. Dan kun je nog een business vault hebben, daar beginnen de business rules te komen. Dan ga je naar de data mart, daar worden vaak heel veel van de gegevens aan elkaar geknoopt. Dat is veel complexer. Dus hoe meer je meeneemt naar boven toe, hoe complexer het verhaal wordt. Dan zie je ook dat je dat niet in 1 keer goed kunt opleveren. Wat ik gelezen heb, hoe dat in theorie zou moeten lopen, is dat je je echt moet beperken. Dat je dan maar die 2 veldjes gaat opleveren. En dan maar op de koop moet toenemen, dat je later nog weer extra werk moet doen om die makkelijke veldjes nog toe te voegen.” (Transcript-Tester, Paragraph 64)

Code	Text
Development method	“Dat planningsproces is nog niet heel erg zorgvuldig. Het lastige van plannen is, dat je ongeveer moet weten hoe ingewikkeld iets is. Er speelt nogal eens onderschatting van hoeveel tijd iets kost.” (Transcript-Tester, Paragraph 50) “Ja. Er worden van te voren refinement sessies gedaan. De materie wordt van te voren besproken. De informatie analysten zijn vaak in de vorige sprint, met de materie voor de volgende sprint. Die zijn dan al aan het onderzoeken waar het vandaan komt. Proberen een beeld te krijgen van de complexiteit. Dat wordt vaak nog niet tot in detail besproken met de ontwikkelaars. Die komen dan later pas erachter dat het een stuk complexer is dan verwacht. Er komen vragen vanuit de ontwikkelaars, naar de informatie analysten. Die moeten dat vaak weer aan de materie deskundige gaan vragen. Die man is er niet altijd, omdat die niet embedded is. Daar zie je al dat het niet lekker loopt.” (Transcript-Tester, Paragraph 52) “Nou is dat vaak in BI land heel lastig te doen, omdat je met een DWH te maken hebt, met verschillende lagen. Dat zie je ook in Data Science. Daar heb je allemaal stappen. Die stappen zitten anders in elkaar, maar ik zie daar parallellen. Je hebt de eerste laag dat gegevens overgehaald worden uit bronsystemen. Of ingelezen worden vanuit bestanden. Dan komt er een laag waarin historie wordt vast gelegd. Dan komt er een laag waarin gegevens in een vorm worden gepresenteerd die voor gebruikers prettig is om op te querien. Je moet die verschillende lagen door. Wat je vaak ziet, is dat met in 1 sprint het ontsluiten van de gegevens doet en het aanmaken van de historische laag, en dat de volgende laag, de data mart laag, de gebruiksvriendelijke laag. Dat dat pas weer een volgende sprint wordt opgepakt. Dus je levert eigenlijk niet in 1 sprint heel veel business waarde op. Dan zie je dus ook dat de domein deskundige, zit te klagen dat het zo lang duurt voordat hij iets krijgt.” (Transcript-Tester, Paragraph 63)
	“Het belangrijkste is dat je realiseert dat Scrum is geen black box. Het is iets wat je gewoon juist met hoge transparantie, hoge frequentie van oplevering” (Ttranscript-Directeur 2-Principle Expert, Paragraph 57) “ik denk dat het jou als opdrachtgever bij uitstek alle mogelijkheden geeft om het project bij te sturen. Daar heb ik hier ook iets over gezegd, moet je maar even lezen.” (Ttranscript-Directeur 2-Principle Expert, Paragraph 57)
Code	Text
Use of Agile Best	“Data in kunnen tekenen, data leken.” (Transcript- Directeur 2-Principle Expert, Paragraph 11)
	“continuous deployment” (Interview transcript- BI consultant, Paragraph 118) “refinement” (Interview transcript- BI consultant, Paragraph 233)
Code	Text
Transition classic to Agile	“Dus naast bij de klant betrokken of actief zijn, had je ook nog coaching van medewerkers, of opleiding of een commerciële taak. Zo werden de taken verdeeld in de groep door ervaren mensen.” (Transcript- Directeur 1- Principal Manager, Paragraph 30)
	“trainingen aangeboden aan de medewerkers om het model,” (Transcript- Directeur 1- Principal Manager, Paragraph 31)

	“Er zijn eerst begonnen met een selecte groep. De meer ervaren mensen. Uiteindelijk werden dat de principles. En die zijn allemaal op een common sense training gegaan, zoals dat heet. Dat was de voorloper van het agile model.” (Transcript- Directeur 1- Principal Manager, Paragraph 32) “Verder dialoog voeren.” (Transcript- Directeur 1- Principal Manager, Paragraph 36) “We ook heel veel, innovatie initiatieven zijn we mee gestopt. Omdat er eigenlijk geen klantvraag achter zat.” (Transcript- Directeur 1- Principal Manager, Paragraph 40) “Iedere organisatievorm volgens mij, is dat het skelet van je huis, of de fundering van je huis. Zonder zo’n afspraken systeem heb je eigenlijk geen fundament. Dus we moeten een afspraken systeem hebben, om met elkaar om te gaan.” (Transcript- Directeur 1- Principal Manager, Paragraph 47) “Want die managementlaag wil niet loslaten. Daar zit het grote probleem.” (Transcript- Directeur 1- Principal Manager, Paragraph 74) “Ze willen van nature uit in control zijn. Wat wij doen is vanuit vertrouwen acteren en we vertrouwen erop dat mensen naar eer en geweten handelen.” (Transcript- Directeur 1- Principal Manager, Paragraph 74)
Code	Text
Agile team	“Door op basis van gelijkwaardigheid in die organisatie te acteren.” (Transcript- Directeur 1- Principal Manager, Paragraph 53)
	“Vakmanschap, de dingen die je dagelijks doet. Is een basisvoorwaarde.” (Transcript- Directeur 1- Principal Manager, Paragraph 76)
	“Niet elke ontwikkelaar vindt testen een leuke activiteit. Maar ze kunnen het vaak heel goed.” (Transcript-Tester, Paragraph 25)
	“Niet iedereen is inwisselbaar. Een tester kan niet zomaar het werk van een ontwikkelaar overnemen. Soms kan dat wel, als je bijv. met je Java achtergrond als tester ingezet wordt. Dan zou je dingen kunnen overnemen.” (Transcript-Tester, Paragraph 27)
	“Kijkend naar de tester. Vroeger kwam die helemaal aan het eind. Ik zit echt bovenop de ontwikkelaars. Op het moment dat zij een eerste stukje programmatuur opleveren. Heb ik de eerste basis testen, de technische testen, heb ik al klaar staan. Een half uur later kunnen ze al de eerste uitslagen krijgen. Anders lukt het ons niet om iets binnen een sprint af te krijgen.” (Transcript-Tester, Paragraph 29)
	“Ik kijk ook heel erg naar de documentatie. Ik heb goede contacten met de informatieanalist. Want als de testbasis onduidelijk is, dan is het moeilijk te testen. Ik moet die kennis gaan halen. Dus daar zit ik ook bovenop. Ontwikkelaars helpen mee met testen. Wat heel belangrijk is binnen agile is dat je de volledigheid test. Worden gegevens goed doorgelaten van laag A naar laag B. Verdwijnt er niets onderweg. Dat zijn vaak ingewikkelde scripts. Die maken de ontwikkelaars vaak ook. Maar ik pas ze ook aan. Ik schrijf ook stukken. Dus dat is echt samenwerken.” (Transcript-Tester, Paragraph 31)
	“Wij hebben een team van ongeveer 6 personen. We hebben de Scrum Master, die ook voor een deel informatieanalist is, dan hebben we een aparte informatieanalist. Omdat we vaak met meerdere trajecten bezig zijn binnen een sprint, verdelen zij het werk en kunnen ze elkaars werk overnemen. Dan hebben we 3 ontwikkelaars. Die elke een eigen specialisme hebben. We werken met PowerCenter, Informatica PowerCenter. Ik ben de BI tester en dat is het team. We hebben nog, en dat is 1 van de problemen, we hebben de business kant. We hebben een materie deskundige. Maar die is niet heel erg embedded in

	het team. Dat heb ik bij meerdere agile teams gezien, dat het vaak niet zo is. En dat is vaak het probleem. Gisteren was er een mooi voorbeeld. Dan blijken de specs toch niet duidelijk genoeg te zijn en dan gaat de man vanaf vandaag weer 2 weken met vakantie. Hij had een college, maar die is onlangs vertrokken, dus er is een nieuwe man en daar kun je nog niet heel veel. Dat is een impediment.” (Transcript-Tester, Paragraph 33)
	“PO levert onderwerpen aan. Die worden niet altijd heel keurig in de klassieke user story vorm ‘als … wil ik …’” (Transcript-Tester, Paragraph 40) “Product owner natuurlijk is een belangrijke.” (Transcript- Directeur 2-Principle Expert, Paragraph 26)
	“Ja, we hebben ook nog een PO. Die weet ook wel het 1 en ander, maar bepaalde stukken niet in details. Dat is echt voor de SME, de materie deskundige. We hebben ook een PO die de prioriteiten stelt.” (Transcript-Tester, Paragraph 36)
	“Maar ons team is ook een beetje apart, omdat een deel, dat zijn meer de technisch georiënteerde mensen, dus de 3 ontwikkelaars en ik als tester, wij horen bij ICT. Terwijl de scrum master, informatie analisten en de SME horen bij de business. De Product Owner hoort ook bij de business.” (Transcript-Tester, Paragraph 42)
	“En je hebt de scrum master, dat zijn denk ik wel de twee. En daaronder krijg je natuurlijk allerlei rollen die verdeeld worden over de taken heen. En de product owner, eigenlijk moet dat de klant zijn. En de scrum master dat zullen wij dan moeten zijn. Zo vullen we de rollen meestal in.” (Transcript- Directeur 2-Principle Expert, Paragraph 26)
	“data engineers, daar zullen data scientists in zitten, er zullen mensen in zitten die op een andere manier werken dan dat je gaat inzetten van, hé, ik heb hier een groep die dan dtl is, en hier heb ik een groep die het via data full doet. Het is een ander kennisgebied.” (Transcript- Directeur 2-Principle Expert, Paragraph 27)
	“Ze willen van nature uit in control zijn. Wat wij doen is vanuit vertrouwen acteren en we vertrouwen erop dat mensen naar eer en geweten handelen.” (Transcript- Directeur 1-Principal Manager, Paragraph 74)
	“Het komt een keer bij elkaar. Het is wel goed, aan het begin begonnen we er over, moet er een architect meekijken. Het is al wel goed om die data scientist, of dat soort jongens die met die data spelen hoe je ze ook noemt, een beetje het gevoel hebben in wat voor soort architectuurwereld leven wij nu. Ze zullen ook toegang moeten hebben tot data die ergens staat. Dus we moeten wel praten met elkaar, maar het moet niet een beperking zijn. Want hun creativiteit ligt heel sterk in het feit dat zij natuurlijk op een heel andere manier denken. Dat zij de data gewoon intanken in een noSQL database en dan maar eens kijken wat er gebeurt, terwijl die andere jongens zitten gewoon in een data Scylla database waar het meestal zit.” (Transcript- Directeur 2-Principle Expert, Paragraph 29)
	“Het zit niet zozeer in dat ze dat niet willen, maar het zijn heel andere mensen. Die data scientists zijn helemaal in een vereende wereld, die zijn helemaal los van alles en die bedenken, maar die anderen moeten het helemaal in een structuur.” (Transcript- Directeur 2-Principle Expert, Paragraph 30)

	“En als je weet waar iedereen zijn kwaliteiten zitten, als je weet hoe je elkaars kwaliteiten kan gaan gebruiken, dan kan je uiteindelijk veel verder komen.” (Transcript- Directeur 2- Principle Expert, Paragraph 40)
	“Domain expert of business analyst” (Interview transcript- BI consultant, Paragraph 14)
	“De specialisten, de ETLers” (Interview transcript- BI consultant, Paragraph 19)
	“je hebt toch vaak verschillende karakters voor nodig.” (Interview transcript- BI consultant, Paragraph 108)
	“Daar hebben we wel een scheiding tussen het ontwikkelwerk en aan de voorkant” (Interview transcript- BI consultant, Paragraph 110)
	“Dus dat Scrum team, vind ik altijd prettig tussen 8-12 personen” (Interview transcript- BI consultant, Paragraph 110) “Agile coach” (Interview transcript- BI consultant, Paragraph 266)
Code	Text
Agile organisation	“wij geloven dat het gedrag van de mensen in een team of een organisatie, bepaald wordt door de organisatie zelf weer.” (Transcript- Directeur 1- Principal Manager, Paragraph 7)
	“enige organisatie inricht zullen mensen zich gaan gedragen naar het ontwerp van die lenige organisatie.” (Transcript- Directeur 1- Principal Manager, Paragraph 8)
	“structuur vasthoudt aan de waarde die je vaststelt met elkaar.” (Transcript- Directeur 1- Principal Manager, Paragraph 8)
	“betrokken bij de ontwikkeling van de organisatie.” (Transcript- Directeur 1- Principal Manager, Paragraph 23)
	“Het verschil met klassiek t.o.v. Agile is dat je die afspraken samen maakt?” (Transcript- Directeur 1- Principal Manager, Paragraph 50)
	“continuous improvement.” (Transcript- Directeur 1- Principal Manager, Paragraph 51)
	“Door op basis van gelijkwaardigheid in die organisatie te acteren.” (Transcript- Directeur 1- Principal Manager, Paragraph 52) “kennen geen hiërarchie in de organisatie.” (Transcript- Directeur 1- Principal Manager, Paragraph 52)
	“Het loslaten van de macht.” (Transcript- Directeur 1- Principal Manager, Paragraph 58)
	“Waar je naar toe wil is een eco systeem waarbij iedereen, met elkaar in verbinding staat op wat voor manier dan ook” (Transcript- Directeur 1- Principal Manager, Paragraph 62)
	“Zij zijn het gedrag gaan vertonen wat wij faciliteren wat vanuit de filosofie beschreven is.” (Transcript- Directeur 1- Principal Manager, Paragraph 77) “Daar hebben we wel een scheiding tussen het ontwikkelwerk en aan de voorkant” (Interview transcript- BI consultant, Paragraph 110)

Code	Text
Use of development method	“Dat planningsproces is nog niet heel erg zorgvuldig. Het lastige van plannen is, dat je ongeveer moet weten hoe ingewikkeld iets is. Er speelt nogal eens onderschatting van hoeveel tijd iets kost.” (Transcript-Tester, Paragraph 50)
	“Ja. Er worden van te voren refinement sessies gedaan. De materie wordt van te voren besproken. De informatie analysten zijn vaak in de vorige sprint, met de materie voor de volgende sprint. Die zijn dan al aan het onderzoeken waar het vandaan komt. Proberen een beeld te krijgen van de complexiteit. Dat wordt vaak nog niet tot in detail besproken met de ontwikkelaars. Die komen dan later pas erachter dat het een stuk complexer is dan verwacht. Er komen vragen vanuit de ontwikkelaars, naar de informatie analysten. Die moeten dat vaak weer aan de materie deskundige gaan vragen. Die man is er niet altijd, omdat die niet embedded is. Daar zie je al dat het niet lekker loopt.” (Transcript-Tester, Paragraph 52)
	“Het belangrijkste is dat je realiseert dat Scrum is geen black box. Het is iets wat je gewoon juist met hoge transparantie, hoge frequentie van oplevering” (Transcript-Directeur 2-Principle Expert, Paragraph 57)
	“Nou is dat vaak in BI land heel lastig te doen, omdat je met een DWH te maken hebt, met verschillende lagen. Dat zie je ook in Data Science. Daar heb je allemaal stappen. Die stappen zitten anders in elkaar, maar ik zie daar parallellen. Je hebt de eerste laag dat gegevens overgehaald worden uit bronsystemen. Of ingelezen worden vanuit bestanden. Dan komt er een laag waarin historie wordt vast gelegd. Dan komt er een laag waarin gegevens in een vorm worden gepresenteerd die voor gebruikers prettig is om op te querien. Je moet die verschillende lagen door. Wat je vaak ziet, is dat met in 1 sprint het ontsluiten van de gegevens doet en het aanmaken van de historische laag, en dat de volgende laag, de data mart laag, de gebruiksvriendelijke laag. Dat dat pas weer een volgende sprint wordt opgepakt. Dus je levert eigenlijk niet in 1 sprint heel veel business waarde op. Dan zie je dus ook dat de domein deskundige, zit te klagen dat het zo lang duurt voordat hij iets krijgt.” (Transcript-Tester, Paragraph 63)
	“er moet 1 bron ontsluiten. Wil je business waarde opleveren? Dan zou je wellicht 1 of 2 veldjes bijv. moeten oppakken voor de 1e laag, die moet doorzetten naar de volgende laag, de data vault laag bijv. En die weer moeten doorzetten naar de data mart laag. Dus dan ga je heel smal opereren. Dat is iets wat een beetje haaks staat op de oude praktijk binnen BI. Daarin ontsluit je de bron en dan neem je zoveel mogelijk gegevens mee die je eventueel later nog kan gebruiken. Je hoort altijd die ontwikkelaars zeggen, die velden hebben we nu eigenlijk nog niet nodig, maar we nemen ze mee want dat is wel zo makkelijk. De eerste laag is dat nog niet zo erg. Dat is redelijk 1 op 1, maar bij de volgende laag begint dat veel complexer te worden. En bij de daarop volgende laag is het nog veel complexer, want bij een DWH neemt de complexiteit van de gegevens, van de business rules toe, naarmate je verder in de pyramide van de gegevens gaat. Want de 1e laag heeft nauwelijks business rules, vrijwel nooit. De 2e laag, de raw vault van de data vault, is redelijk simpel. Dan kun je nog een business vault hebben, daar beginnen de business rules te komen. Dan ga je naar de data mart, daar worden vaak heel veel van de gegevens aan elkaar geknoopt. Dat is veel complexer. Dus hoe meer je meeneemt naar boven toe, hoe complexer het verhaal wordt. Dan zie je ook dat je dat niet in 1 keer goed kunt opleveren. Wat ik gelezen heb, hoe dat in theorie zou moeten lopen, is dat je je echt moet

	beperken. Dat je dan maar die 2 veldjes gaat opleveren. En dan maar op de koop moet toenemen, dat je later nog weer extra werk moet doen om die makkelijke veldjes nog toe te voegen." (Transcript-Tester, Paragraph 64)
	"ik denk dat het jou als opdrachtgever bij uitstek alle mogelijkheden geeft om het project bij te sturen. Daar heb ik hier ook iets over gezegd, moet je maar even lezen." (Transcript-Directeur 2-Principle Expert, Paragraph 57)
Code	Text
Transition classic to Agile	"We hadden namelijk voor iedere, nou ja we hadden veel management. Dus een consultant mocht zich alleen maar bezighouden met de opdracht die hij kreeg. Hij mocht niet meepraten, niet meedenken, heel gechargeerd, over allerlei nieuwe ontwikkelingen. Want dat was het management dat zich daarmee bemoeide. Ze hadden zelf geen zeggenschap over het tarief waarvoor ze ingezet werden. Ze konden dus ook geen invloed uitoefenen op de waarde die ze toevoegde voor de klant en hetgeen Ordina daarvoor terugkreeg. Dat soort aspecten, maakt, dat als een consultant weinig invloed heeft, hij zich ook zal terugtrekken. Als hij 2 keer of 3 keer toe een poging gedaan heeft om invloed te hebben, en het is weinig succesvol, dan stopt hij daarmee." (Transcript- Directeur 1-Principal Manager, Paragraph 14)
	"Vervolgens zijn we gaan kijken hoe we dat konden zouden kunnen veranderen. Hoe we consultants mee zouden kunnen nemen in de verandering die eraan zat, of die we wilden doorvoeren. Er moest iets veranderen. We maakten veel verlies. We moesten mensen gaan ontslaan, we hadden veel te veel overhead, al die dingen die speelden." (Transcript-Directeur 1- Principal Manager, Paragraph 16)
	"Uiteindelijk zijn er een zevental werkgroepen ontstaan om te kijken hoe we een verandering vorm zouden kunnen geven." (Transcript- Directeur 1- Principal Manager, Paragraph 19)
	"Uiteindelijk hebben we met een groep mensen gekeken wat we zouden kunnen veranderen op basis van de constateringen die we uit de groep gehaald hebben. Daar is uit voren gekomen dat we een plattere organisatie wilden, dat mensen meer invloed wilden, dat ze veel meer betrokken wilden zijn bij het primaire proces. Uiteindelijk hebben we gekeken hoe we dat zouden kunnen realiseren. Een van de modellen die toen naar voren kwam was het Spotify model." (Transcript- Directeur 1- Principal Manager, Paragraph 24)
	"We hebben gesaneerd, alle manager rollen zijn verdwenen. Iedereen heeft een directe rol gekregen in de organisatie. En afhankelijk van senioriteit of rol had je een aantal neven taken." (Transcript- Directeur 1- Principal Manager, Paragraph 29)
	"Dus naast bij de klant betrokken of actief zijn, had je ook nog coaching van medewerkers, of opleiding of een commerciële taak. Zo werden de taken verdeeld in de groep door ervaren mensen." (Transcript- Directeur 1- Principal Manager, Paragraph 30)
	"Trainingen aangeboden aan de medewerkers om het model," (Transcript- Directeur 1-Principal Manager, Paragraph 31)
	"Er zijn eerst begonnen met een selecte groep. De meer ervaren mensen. Uiteindelijk werden dat de principles. En die zijn allemaal op een common sense training gegaan, zoals dat heet. Dat was de voorloper van het agile model." (Transcript- Directeur 1- Principal Manager, Paragraph 32)

	“Verder dialoog voeren.” (Transcript- Directeur 1- Principal Manager, Paragraph 36)
	“We ook heel veel, innovatie initiatieven zijn we mee gestopt. Omdat er eigenlijk geen klantvraag achter zat.” (Transcript- Directeur 1- Principal Manager, Paragraph 40)
	“Iedere organisatievorm volgens mij, is dat het skelet van je huis, of de fundering van je huis. Zonder zo’n afspraken systeem heb je eigenlijk geen fundament. Dus we moeten een afspraken systeem hebben, om met elkaar om te gaan.” (Transcript- Directeur 1- Principal Manager, Paragraph 47)
	“Want die managementlaag wil niet loslaten. Daar zit het grote probleem.” (Transcript- Directeur 1- Principal Manager, Paragraph 74)
	“Ze willen van nature uit in control zijn. Wat wij doen is vanuit vertrouwen acteren en we vertrouwen erop dat mensen naar eer en geweten handelen.” (Transcript- Directeur 1- Principal Manager, Paragraph 74)
Code	Text
Agile team	“Door op basis van gelijkwaardigheid in die organisatie te acteren.” (Transcript- Directeur 1- Principal Manager, Paragraph 53)
	“Vakmanschap, de dingen die je dagelijks doet. Is een basisvoorwaarde.” (Transcript- Directeur 1- Principal Manager, Paragraph 76) “Managers willen van nature uit in control zijn. Wat wij doen is vanuit vertrouwen acteren en we vertrouwen erop dat mensen naar eer en geweten handelen.” (Transcript- Directeur 1- Principal Manager, Paragraph 74)

	“En je hebt de scrum master, dat zijn denk ik wel de twee. En daaronder krijg je natuurlijk allerlei rollen die verdeeld worden over de taken heen. En de product owner, eigenlijk moet dat de klant zijn. En de scrum master dat zullen wij dan moeten zijn. Zo vullen we de rollen meestal in.” (Transcript- Directeur 2-Principle Expert, Paragraph 26) “Data engineers, daar zullen data scientists in zitten, er zullen mensen in zitten die op een andere manier werken dan dat je gaat inzetten van, hé, ik heb hier een groep die dan dtl is, en hier heb ik een groep die het via data full doet. Het is een ander kennisgebied.” (Transcript- Directeur 2-Principle Expert, Paragraph 27) “En als je weet waar iedereen zijn kwaliteiten zitten, als je weet hoe je elkaars kwaliteiten kan gaan gebruiken, dan kan je uiteindelijk veel verder komen.” (Transcript- Directeur 2-Principle Expert, Paragraph 40) “Het komt een keer bij elkaar. Het is wel goed, aan het begin begonnen we er over, moet er een architect meekijken. Het is al wel goed om die data scientist, of dat soort jongens die met die data spelen hoe je ze ook noemt, een beetje het gevoel hebben in wat voor soort architectuurwereld leven wij nu. Ze zullen ook toegang moeten hebben tot data die ergens staat. Dus we moeten wel praten met elkaar, maar het moet niet een beperking zijn. Want hun creativiteit ligt heel sterk in het feit dat zij natuurlijk op een heel andere manier denken. Dat zij de data gewoon intanken in een noSQL database en dan maar eens kijken wat er gebeurt, terwijl die andere jongens zitten gewoon in een data Scylla database waar het meestal zit.” (Transcript- Directeur 2-Principle Expert, Paragraph 29)
	“Domein expert of business analyst” (Interview transcript- BI consultant, Paragraph 14) “De specialisten, de ETLers” (Interview transcript- BI-consultant, Paragraph 19) “je hebt toch vaak verschillende karakters voor nodig.” (Interview transcript- BI consultant, Paragraph 108) “Dus dat Scrum team, vind ik altijd prettig tussen 8-12 personen” (Interview transcript- BI-consultant, Paragraph 110) “Agile coach” (Interview transcript- BI consultant, Paragraph 266)
	“Niet elke ontwikkelaar vindt testen een leuke activiteit. Maar ze kunnen het vaak heel goed.” (Transcript-Tester, Paragraph 25) “Niet iedereen is inwisselbaar. Een tester kan niet zomaar het werk van een ontwikkelaar overnemen. Soms kan dat wel, als je bijv. met je Java achtergrond als tester ingezet wordt. Dan zou je dingen kunnen overnemen.” (Transcript-Tester, Paragraph 27) “Kijkend naar de tester. Vroeger kwam die helemaal aan het eind. Ik zit echt bovenop de ontwikkelaars. Op het moment dat zij een eerste stukje programmatuur opleveren.” (Transcript-Tester, Paragraph 29) “PO levert onderwerpen aan. Die worden niet altijd heel keurig in de klassieke user story vorm ‘als … wil ik …’” (Transcript-Tester, Paragraph 40)

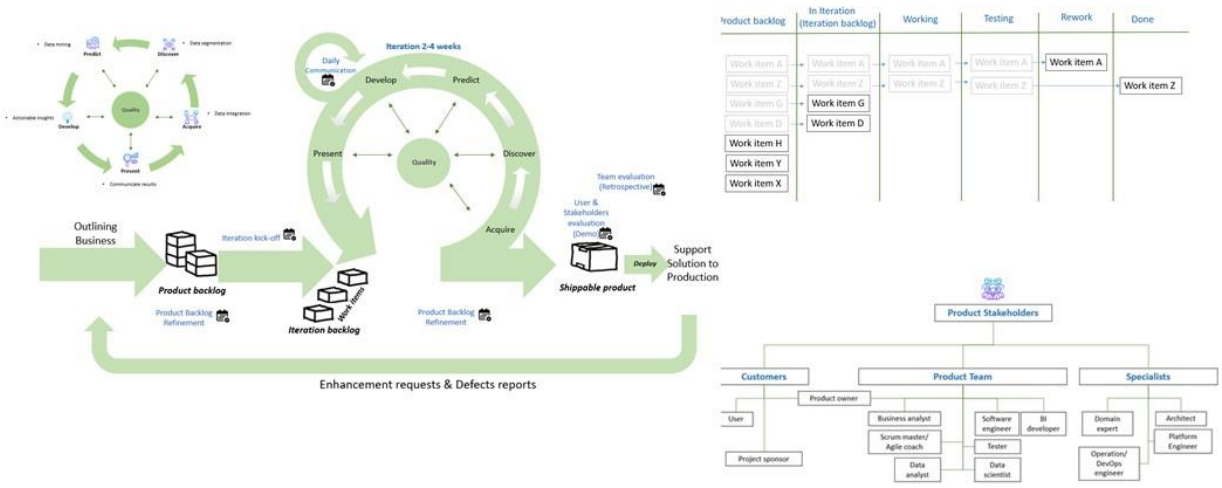
	“Product owner natuurlijk is een belangrijke.” (Transcript- Directeur 2-Principle Expert, Paragraph 26) “Ja, we hebben ook nog een PO. Die weet ook wel het 1 en ander, maar bepaalde stukken niet in details. Dat is echt voor de SME, de materie deskundige. We hebben ook een PO die de prioriteiten stelt.” (Transcript-Tester, Paragraph 36) “Maar ons team is ook een beetje apart, omdat een deel, dat zijn meer de technisch georiënteerde mensen, dus de 3 ontwikkelaars en ik als tester, wij horen bij ICT. Terwijl de scrum master, informatieanalisten en de SME horen bij de business. De Product Owner hoort ook bij de business.” (Transcript-Tester, Paragraph 42)
Code	Text
Agile organisation	“Wij geloven dat het gedrag van de mensen in een team of een organisatie, bepaald wordt door de organisatie zelf weer.” (Transcript- Directeur 1- Principal Manager, Paragraph 7) “Enige organisatie inricht zullen mensen zich gaan gedragen naar het ontwerp van die lenige organisatie.” (Transcript- Directeur 1- Principal Manager, Paragraph 8) “Structuur vasthoudt aan de waarde die je vaststelt met elkaar.” (Transcript- Directeur 1- Principal Manager, Paragraph 8) “Betrokken bij de ontwikkeling van de organisatie.” (Transcript- Directeur 1- Principal Manager, Paragraph 23) “Het verschil met klassiek t.o.v. Agile is dat je die afspraken samen maakt?” (Transcript- Directeur 1- Principal Manager, Paragraph 50) “continuous improvement.” (Transcript- Directeur 1- Principal Manager, Paragraph 51) “Door op basis van gelijkwaardigheid in die organisatie te acteren.” (Transcript- Directeur 1- Principal Manager, Paragraph 52) “Kennen geen hiërarchie in de organisatie.” (Transcript- Directeur 1- Principal Manager, Paragraph 52) “Het loslaten van de macht.” (Transcript- Directeur 1- Principal Manager, Paragraph 58) “Waar je naar toe wil is een eco systeem waarbij iedereen, met elkaar in verbinding staat op wat voor manier dan ook” (Transcript- Directeur 1- Principal Manager, Paragraph 62)
	“Zij zijn het gedrag gaan vertonen wat wij faciliteren wat vanuit de filosofie beschreven is.” (Transcript- Directeur 1- Principal Manager, Paragraph 77) “Daar hebben we wel een scheiding tussen het ontwikkelwerk en aan de voorkant” (Interview transcript- BI-consultant, Paragraph 110)

Appendix VI- Evaluation survey

Thank you very much for participating in the evaluation event.

Please fill this survey and let me know your thoughts about the AFDS framework.

*Required



Goal

The goal is characterized by the "Efficacy" which is the degree to which AFDS provides its desired effect based on the requirements and the "Generality" which provides the framework to be applied in a broader project area such as data science, business intelligence etc.
Please read through the following statements and indicate your opinion with each.

1. **Would you introduce AFDS to your client?** * *Mark only one oval.*

	1	2	3	4	5	
Definitely not	◯	◯	◯	◯	◯	Definitely

2. **Can AFDS improve Agile implementation at your project?** * *Mark only one oval.*

	1	2	3	4	5	
Definitely not	◯	◯	◯	◯	◯	Definitely

3 **Can AFDS be applied to any of the following areas?** * *Mark only one oval per row.*

	Definitely not	Probably not	Possibly	Probably	Definitely
Data science	◯	◯	◯	◯	◯
Business Intelligent	◯	◯	◯	◯	◯
Big Data	◯	◯	◯	◯	◯
Linked Data	◯	◯	◯	◯	◯

Environment

People and the organization are chosen as evaluation criteria from the environment dimension to verify the consistency of the framework. 'Understandability', 'ease of use' and 'fit within the organization' are the sub criteria of this evaluation.

Please read through the following statements and decide how much you either agree or disagree with each.

4. **The elements of AFDS framework are easy to understand by all disciplines in a project.** * *Mark only one oval per row.*

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
AFDS Development Cycle					
AFDS Development Method					
AFDS Agile Practices					

5. **The elements of AFDS framework are easy to use.** * *Mark only one oval per row.*

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
AFDS Development Cycle					
AFDS Development Method					
AFDS Agile Practices					

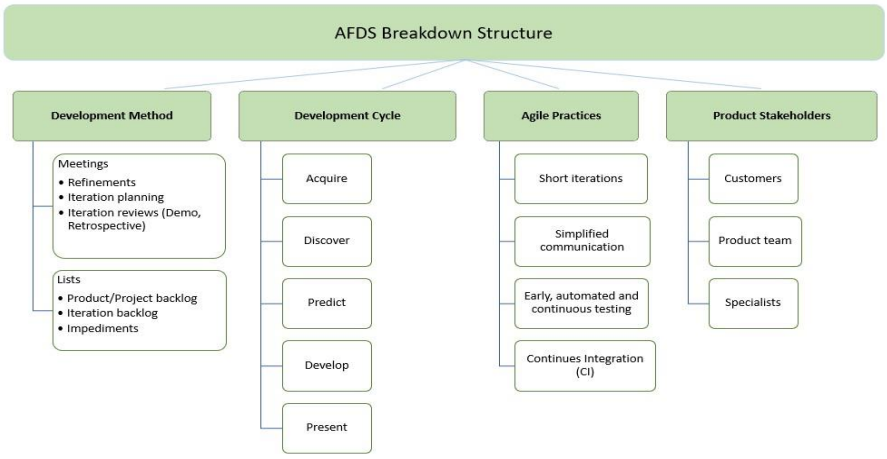
6. **AFDS framework fits within your client's organization.** * *Mark only one oval per row.*

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
AFDS fit in Data Science projects					
ADFS fits in Business Intelligent projects					
ADFS fits in Data Link projects					
AFDS fits in Data Driven projects					

Structure

The structure of AFDS is evaluated by completeness, simplicity, level of detail, and consistency.

Please read through the following statements and decide how much you either agree or disagree or are satisfied or dissatisfied with each.



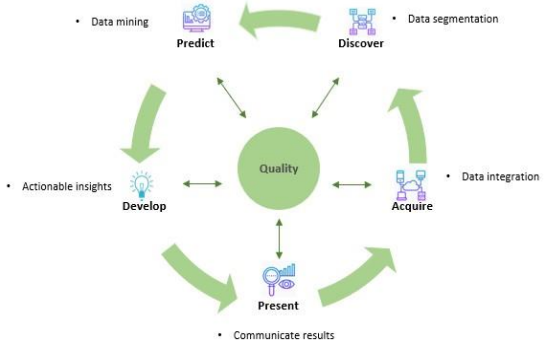
7. **AFDS framework is ...** * *Mark only one oval per row.*

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
Complete					
Easy to apply					

8. **AFDS framework gives enough guidelines about how to implement and apply Agile.** * *Mark only one oval.*

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

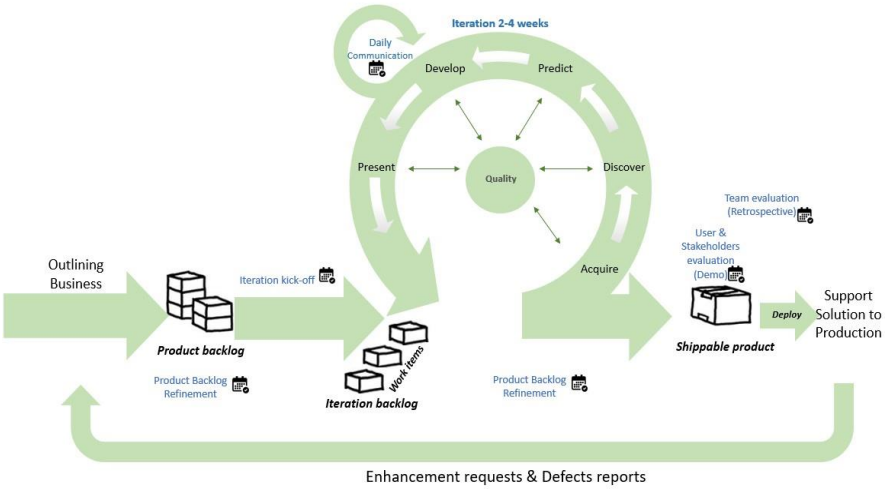
9 **How satisfied are you with the activities of AFDS data development cycle?** *



Mark only one oval per row.

	Very Dissatisfied	Dissatisfied	Natural	Satisfied	Very Satisfied
Acquire - First activity					
Discover - Second activity					
Predict - Third activity					
Develop - Fourth activity					
Present - Fifth activity					

10 **How satisfied are you with the AFDS development method? ***



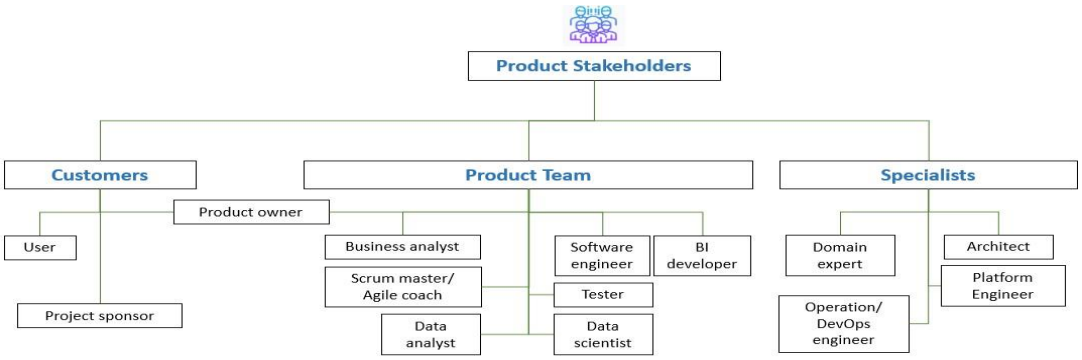
Mark only one oval per row.

	Very Dissatisfied	Dissatisfied	Natural	Satisfied	Very Satisfied
Iterative & incremental process					
Based on Scrum					
Daily communication of team members					
Definition of A Minimum Viable Product (Unit of Work)					
Business Involvement in development process					
Operation Involvement in development process					

11. **How satisfied are you with the AFDS Agile best practices? *** *Mark only one oval per row.*

	Very Dissatisfied	Dissatisfied	Natural	Satisfied	Very Satisfied
Continues improvement					
Continues Integration					
Test Driven Data Development					
Product team definition					
Project stakeholders & role definition					

12 **How satisfied are you with the definition of AFDS product stakeholders? ***



Mark only one oval per row.

	Very dissatisfied	Dissatisfied	Natural	Satisfied	Very satisfied
User					
Project sponsor					
Product Owner					
Business analyst					
Scrum master/Agile coach					
Data analyst					
Software engineer					
Tester					
BI developer					
Data scientist					
Domain expert					
Architect					
Operation / DevOps engineer					
Platform engineer					

Evolution

Learning capability from the evolution dimension is used to evaluate the capacity of the framework from the experience and reaction of the environment.
Please read through the following statements and indicate your opinion.

13. **How relevant and helpful can AFDS be for your team?** *
Mark only one oval.

	1	2	3	4	5	
Not relevant						Very relevant

14. **How satisfied are you with AFDS Framework?** * *Mark only one oval.*

	1	2	3	4	5	
Very dissatisfied						Very setisfied

15 **Do you have any additional comments regarding the framework, evaluation session and the research?**

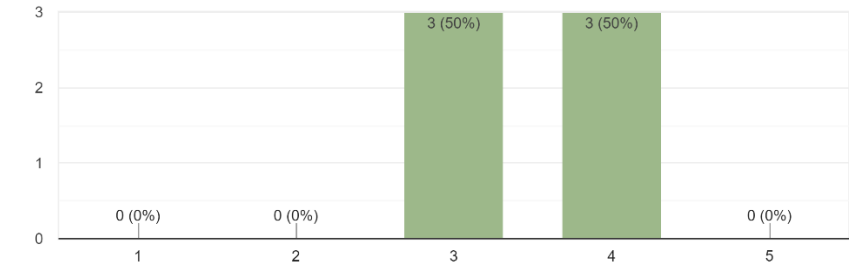
You have reached the end of the questionnaire. Thank you very much for participating in this research.

Appendix VII- Results of the evaluation survey

I. Goal

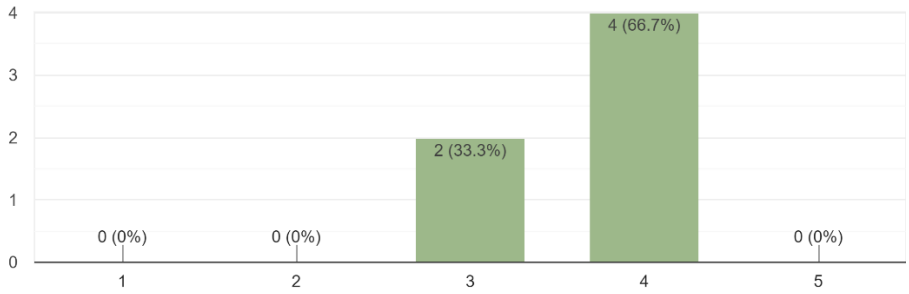
Would you introduce AFDS to your client?

6 responses

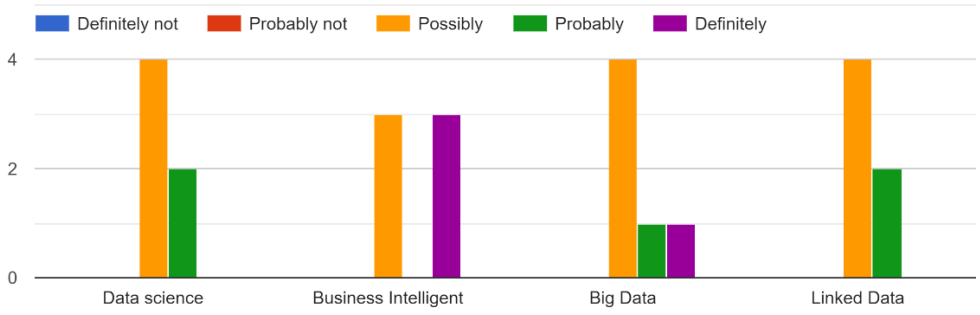


Can AFDS improve Agile implementation at your project?

6 responses

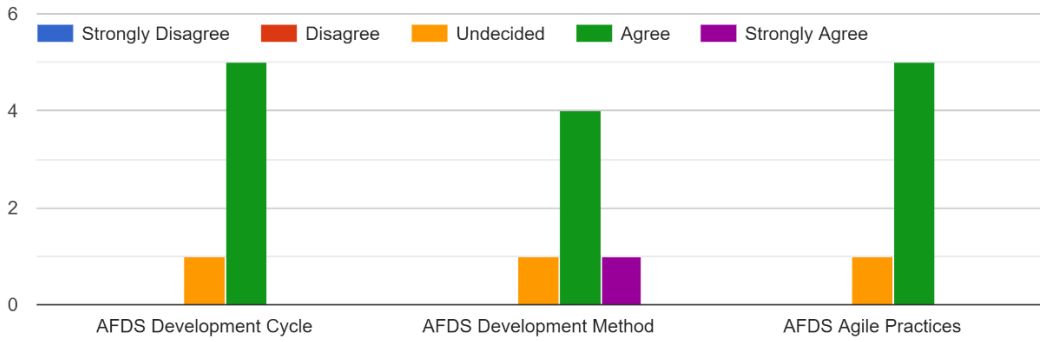


Can AFDS be applied to any of the following areas?

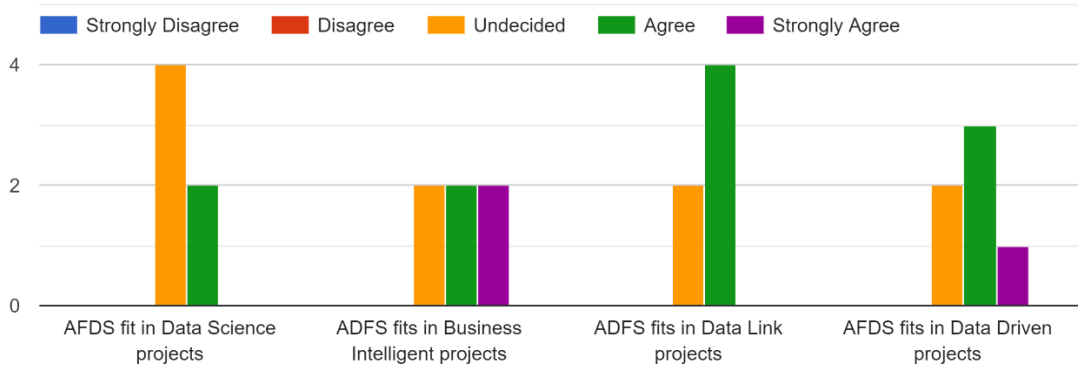


II. Environment

The elements of AFDS framework are easy to understand by all disciplines in a project.

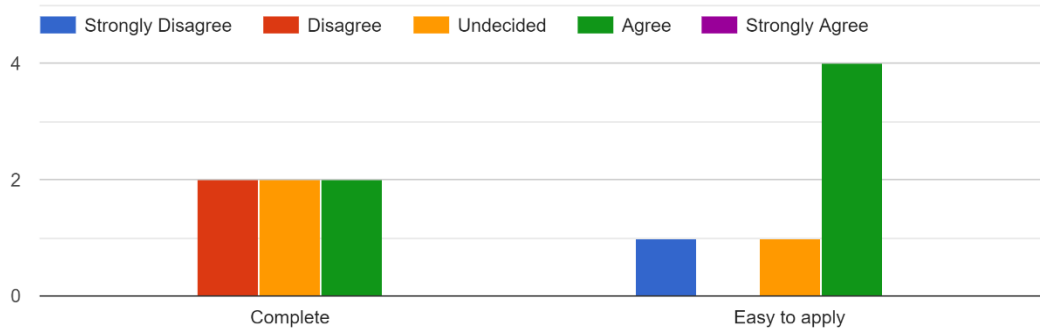


AFDS framework fits within your client's organization.



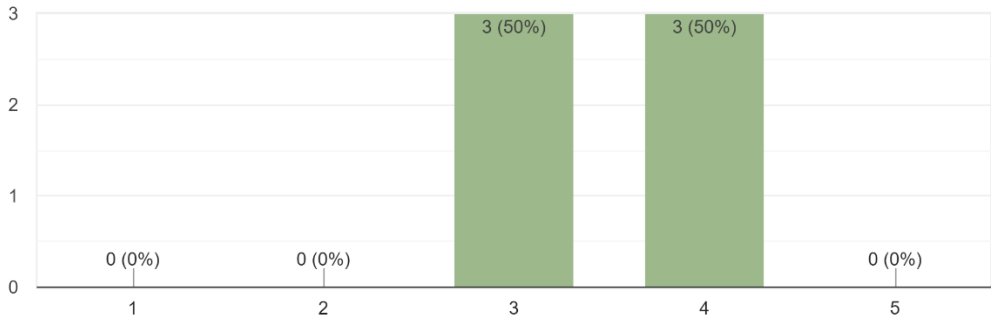
III- Structure

AFDS framework is ...

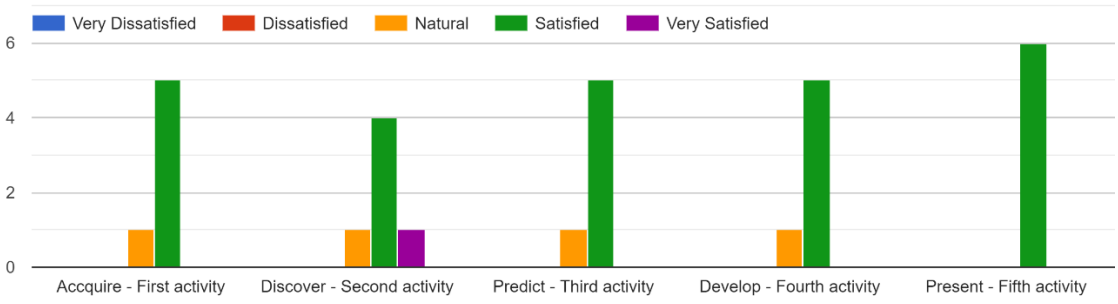


AFDS framework gives enough guidelines about how to implement and apply Agile.

6 responses

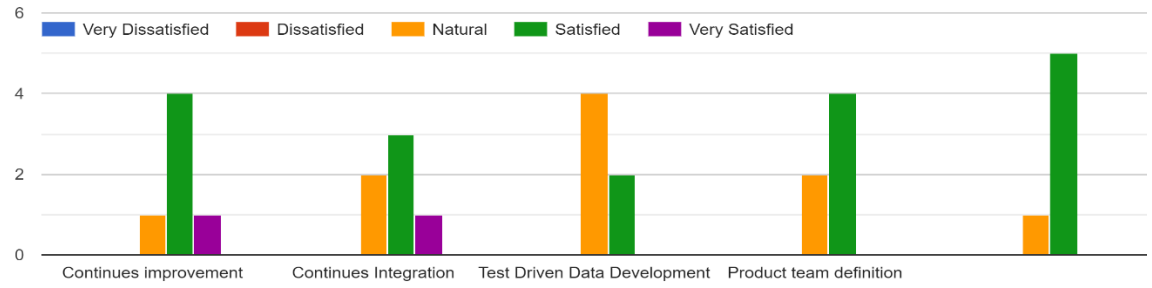


How satisfied are you with the activities of AFDS data development cycle?

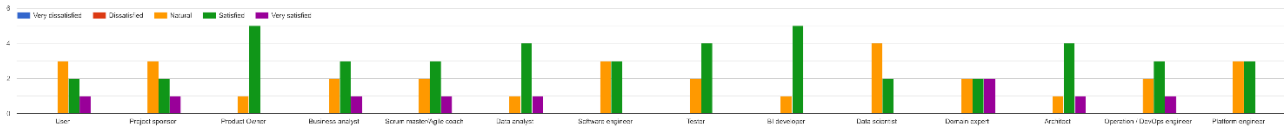


How satisfied are you with the AFDS development method?

How satisfied are you with the AFDS Agile best practices?



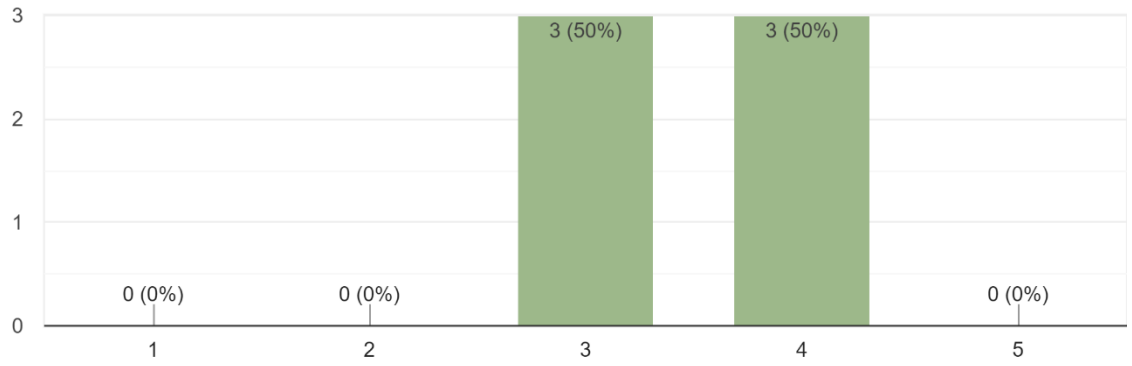
How satisfied are you with the definition of AFDS product stakeholders?



IV- Evolution

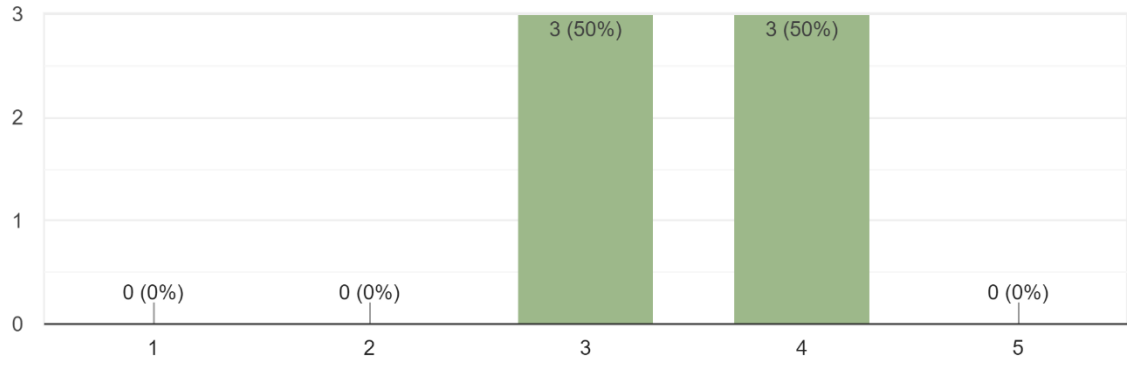
How relevant and helpful can AFDS be for your team?

6 responses



How satisfied are you with AFDS Framework ?

6 responses



Do you have any additional comments regarding the framework, evaluation session and the research?1 response

In addition, I would like a discussion on how to determine MVP's in Data Science projects. I would also like to discuss how this product end up in the AFDS cycle, e.g. back logs.

Appendix VIII- AFDS first concept

In this appendix the first design of the elements is included. These are validated by the first iteration during the validation presentation. Based on the feedback of the experts the elements are modified.

First design of AFDS data development cycle

Figure 23: First design of AFDS data development cycle

First design of AFDS Method

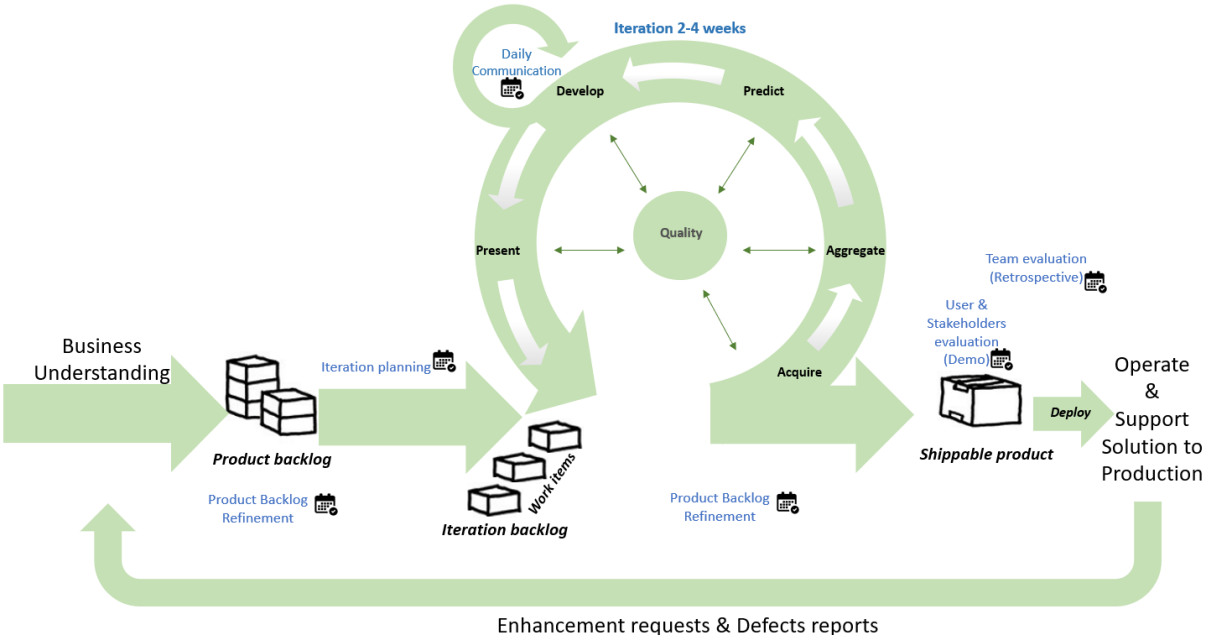


Figure 24: First design of AFDS development method

First design of AFDS project stakeholders

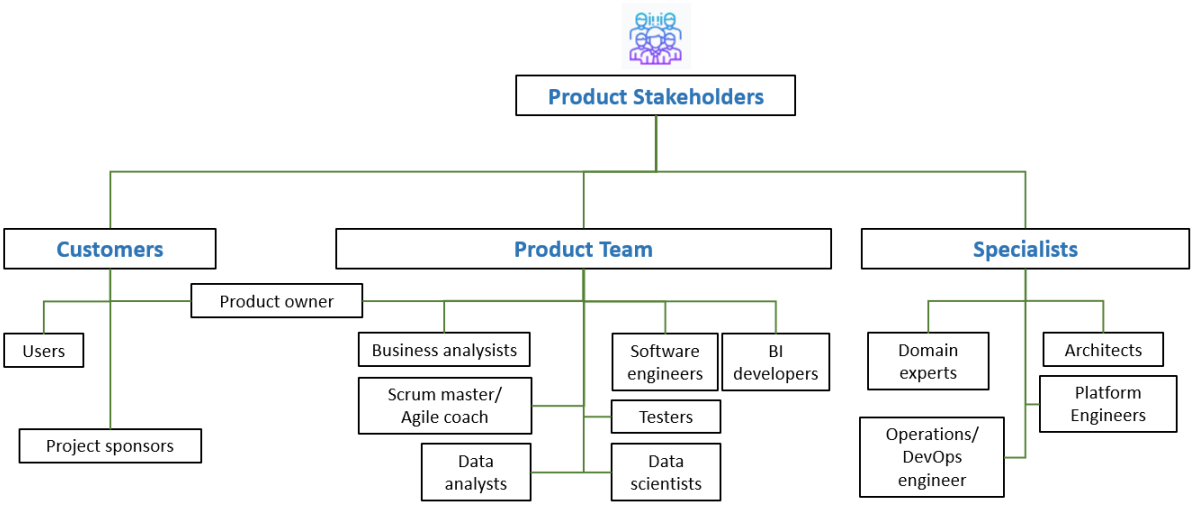


Figure 25: First design of AFDS project stakeholders